بناء اختبار مهارات التفكير الناقد باستخدام نظرية استجابة الفقرة

إعداد

آمال أحمد الزعبي

بكالوريوس رياضيات، جامعة اليرموك، ١٩٨٣

ماجستير أساليب تدريس رياضيات، جامعة اليرموك، ١٩٨٧

قدمت هذه الأطروحة استكمالاً لمتطلبات الحصول على درجة الدكتوراه في تخصص القياس والتقويم في جامعة اليرموك، اربد، الأردن

وافق عليها

أ.د. يوسف محمد سوالمة .... رئيساً

أستاذ القياس والإحصاء التربوي، جامعة اليرموك

أ.د. عدنان يوسف العتوم .... عضواً

أستاذ علم النفس الاجتماعي، جامعة اليرموك

أ.د. ضرار محمد جرادات .... عضواً

أستاذ القياس والتقويم التربوي، جامعة اليرموك

أ.د. أحمد سليمان عودة .... عضواً

أستاذ البحث والتقويم التربوي، جامعة اليرموك

د. أحمد يوسف قواسمة .... عضواً

أستاذ مشارك في القياس والتقويم التربوي، جامعة اليرموك

تاريخ تقديم الأطروحة ٢٠٠٦/٥/٣م

الإهـــداء

إلى الذي أدعو الله أن يكون في عليين . . . والدي رحمه الله

إلى التي ما كلّت تمد ظلها علينا . . . فتقرب لنا المسافات

وتسهّل علينا الصعاب . . . وتمدنا بطاقة لا نعرف . .

لها حدوداً . . . إلى نبع الحب والعطاء . . . والدتي أمدّ الله في عمرها . . .

إلى من كان فرحي فرحهم وحزني حزنهم . . . فقاسموني البسمة والدمعة،

وتحملوا معي العناء والمشقة . . أخواتي وإخواني

إلى كل من ساندني وشد أزري . . . وثبّت فيّ العزيمة من الأصدقاء

والأحباب . . .

أهدي جهدي

الباحثة

شكر وتقدير

الحمد لله الذي أعانني على إتمام هذا العمل، ولا يسعني وقد آلت هذه الرسالة إلى الخروج إلى حيز الوجود إلا أن أتقدّم بجزيل شكري وعظيم امتناني إلى صاحب الخلق الرفيع والمواقف العلمية الرائدة أستاذي الفاضل الأستاذ الدكتور يوسف سوالمة لما أملاه من فيض علمه وحُسن معاملته، والذي فتح لي أبواب المعرفة ولم يتوان لحظةً في مد يد العون والمساعدة وتقديم الدعم والمتابعة الحثيثة بكل رحابة صدر على الرغم من أعبائه الكبيرة، فكل حرف من حروف الرسالة تدين له بالفضل والعرفان، فنعم المشرف الذي أقف عاجزةً عن تقديم الشكر اللائق به، فجزاه الله عني كل خير ورعاه . .

ولا يسعني إلا أن أقف إجلالاً وإكباراً إلى منارات المعرفة والعلم أساتذتي الأفاضل: الأستاذ الدكتور عدنان العتوم، والأستاذ الدكتور ضرار جرادات، والأستاذ الدكتور أحمد عودة، والدكتور أحمد قواسمة، الذين شرّفوني بالتفضل مشكورين في مناقشة هذه الأطروحة والفيض عليها من علمهم وإغنائها، والذين لم يبخلوا عليّ بعطاءاتهم ونصحهم وإرشاداتهم القيّمة . .

كما أتقدم بجزيل الشكر إلى زملائي في تربية جرش لما قدموه لي من عون ومساعدة، وإلى الزملاء في تربية اربد الأولى لما قدموه من تعاون في تطبيق الأداة، وإلى جميع مديرات ومديري المدارس التي تم تطبيق الأداة فيها، وإلى الأعزاء الطلبة والطالبات الذين ساعدوا في إنجاز هذا العمل .

وجزيل شكري للأخ الزميل الدكتور محمد فضل الدلابيح على تفضله بمراجعة وتدقيق الرسالة لغوياً .

وتقديري الكبير إلى كل الأخوة والزملاء الذين تعاونوا معي في إخراج هذه الرسالة، جزاهم الله عني خير الجزاء .

الباحثة

فهرس المحتويات

فهرس الجداول

فهرس الأشكال

الشكل	الصفحة

فهرس الملاحق

الملخص

الزعبي، آمال أحمد. بناء اختبار مهارات التفكير الناقد باستخدام نظرية استجابة الفقرة. أطروحة دكتوراة، جامعة اليرموك 2006 (المشرف: أ.د. يوسف محمد سوالمة).

هدفت الدراسة الحالية إلى بناء اختبار مهارات التفكير الناقد باستخدام نظرية استجابة الفقرة. ولتحقيق هذا الهدف تم تعريف التفكير الناقد، وكتابة 228 فقرة من نوع الاختيار من متعدد تغطي مهارات التفكير الناقد الآتية: التحليل، والتفسير، والتقويم، والشرح، والاستدلال، والتبرير الإحصائي والاحتمالي، والتعرف إلى الافتراضات المنصوص وغير المنصوص عليها، والتركيز. ومن خلال عملية التحكيم المتشعبة والتجريب الأولي تم الاحتفاظ بـ 89 فقرة.

طبق الاختبار على 444 طالباً وطالبة في المرحلة الأساسية العليا، و610 طالباً وطالبة في المرحلة الثانوية، و120 طالباً وطالبة في المرحلة الجامعية. وتم التحقق من افتراض أحادية البعد والصدق والثبات للاختبار من خلال العديد من المؤشرات الدالة عليها. وتم التحقق من مطابقة الاستجابات عن فقرات الاختبار لنماذج نظرية استجابة الفقرة. وبعد حذف الأفراد غير المطابقين والفقرات غير المطابقة تبين أن هناك 50 فقرة طابقت النموذج الأحادي المعلمة، و70 فقرة طابقت النموذج ثنائي المعلمة، و82 فقرة طابقت النموذج ثلاثي المعلمة.

تم استخدام برنامجي (SPSS, Bilog-Mg) لتقدير معالم الفقرات وحساب تقديرات قدرات الأفراد وحساب الدقة النسبية والفاعلية النسبية. أظهرت النتائج أن تقديرات القدرة وفق النموذج ثلاثي المعلمة أكثر دقة منها للنموذجين أحادي وثنائي المعلمة خاصة عند مستويات القدرة العليا. كما أن الفاعلية النسبية للنموذج الثلاثي المعلمة أعلى منها للنماذج الأخرى.

واتصفت معالم الفقرات ومعلم القدرة بخاصية اللاتغير في 67% من الحالات للنموذج الأحادي و67% من الحالات للنموذج الثنائي و78% من الحالات للنموذج الثلاثي. وقد حسبت معايير الاختبار المتمثلة بالرتب المئينية لعلامات الاختبار النهائي (82 فقرة) في عدة مجموعات معيارية. وفي ضوء النتائج ونظراً لأن الاختبار بصورته النهائية (82 فقرة) يتمتع بخصائص سيكومترية مقبولة، توصي الباحثة باستخدامه أو أي اختبار فرعي مستل منه لقياس مهارة التفكير الناقد ولأغراض البحث العلمي. وبضرورة استخدام نظرية استجابة الفقرة في بناء اختبار للتفكير الناقد يتناول أنواعاً أخرى من الفقرات مثل الفقرات المقالية والصح والخطأ المتعدد وغيره.

الكلمات المفتاحية:

التفكير الناقد، نظرية استجابة الفقرة، الدقة النسبية، الفاعلية النسبية، اللاتغير.

الفصل الأول

خلفية الدراسة وأهميتها

تمهيد:

يعتبر التفكير أحد المظاهر التي تميّز الكائن البشري عن غيره من المخلوقات، فالتفكير عملية مستمرة يمارسها المرء يومياً، وذلك في أثناء قيام الدماغ بسلسلة من النشاطات العقلية عندما يتعرض لمثير يتم استقباله عن إحدى طرق الحواس الخمس. ويرى بول (Paul, 1998) أن التفكير عبارة عن سعي حثيث للوضوح والدقة والعدالة المستندة إلى العقل، وحماسة للتوصل إلى عمق الأشياء وجذورها، ومقت شديد للتناقض وعدم الاتساق.

والتفكير عملية هادفة، ونشاط عقلي يتيح لنا الاهتمام بحل المسألة، وتفسير البيانات، وعمل القرارات، وفهم الأفكار والمفاهيم (McWhorter & Collins, 1992). ويؤكد بول (Paul, 1998) أن جودة تفكير المرء تؤثر على جودة حياته، ويشير سالدر (Salder, 1993) إلى أنه حتى يتفوق الإنسان لا بد أن تتكون لديه مهارات التفكير الناقد ويتمكن من بعض استراتيجيات تعليم التفكير. ويضيف شالوبا (Chalupa, 1992) إن إحدى طرق إعداد الفرد للمستقبل هي تعليمه كيف يفكر بدلاً من بماذا يفكر.

ويؤكد ديوي (1960) أن التفكير لا ينشأ إلا عندما تكون الأشياء غير محققة وفي موضع شك، حيث يبدأ الإنسان عادة بالتفكير عندما لا يعرف ما الذي يجب عمله بالتحديد في أثناء مواجهته مواقف تتسم بالحيرة والتردد والغموض في فهم عناصر الموقف. ويشبه مكلير (Maclure) (المشار إليه في جروان، 2002 أ) التفكير بعملية التنفس للإنسان، فكما أن

التنفس عملية لازمة لحياة الإنسان، فإن التفكير أشبه ما يكون بنشاط طبيعي لا غنى للإنسان عنه في حياته اليومية. ويفترض دي بونو (1997) بأن التفكير اكتشاف مترو أو متبصر أو متأن للخبرة من أجل التوصل إلى الهدف.

وقد اهتم الفلاسفة من أمثال أرسطو (Aristotle)، ومن بعده لوك (lock) وهوبس (Hobbes) وميلز (Mills) بوصف التفكير باعتباره مجموعة من الصور العقلية المترابطة والمستخلصة من الخبرة الإدراكية للفرد، وبدراسة العقل باعتباره مقراً لعمليات الاستدلال التي يقوم بها الفرد، وكان اهتمام الفلاسفة بدراسة الناتج عن عملية التفكير من خلال تطبيق قوانين المنطق عليه لمعرفة مدى اتساقه مع الحقائق (السيد، 1995؛ Paul، Elder & Bartell, 2004).

فقد ركزّ سقراط منذ (2500) سنة على أهمية طرح الأسئلة المتعمقة المثيرة للتفكير ، وأكدّ على أهمية البحث عن الدليل وفحص الحجج والافتراضات وتحليل المفاهيم. وبعد سقراط أكدّ كل من أفلاطون (Plato) وأرسطو (Aristotle) على أن الأشياء غالباً ما تختلف عما تظهر عليه، وأن العقل المدرّب جيداً يستطيع أن يرى من خلال الأشياء التي تبدو على السطح. وأما هوبس ولوك (Hobbes and Lock) فقد لعبا دوراً هاماً في مجال التفكير حين أشارا إلى أن العقل المفكر يجب أن يكون متفتحاً للتعليم الجديد، وأكدّ هوبس أن كل شيء يجب أن يتم شرحه بالدليل والتبرير في حين رفض لوك التحليل عن طريق الإحساس العام (Common Sense) (Paul, Elder & Bartell, 2004).

وأخذت دراسة التفكير أشكالاً عديدة، فكان لمحاولات علماء مدرسة الجشطالت مساهمات في دراسة التفكير من خلال معرفتهم عن الإدراك، ومن ثم كان تعريفهم للتفكير باعتباره هيكلة العلاقات بين عناصر المشكلة بطريقة جديدة، واحتل مفهوم اتجاه التفكير مركزاً هاماً لعلماء الجشطالت في مجال دراسة التفكير (السيد، 1995؛ مارزانو وآخرون، 2004).

أما المدرسة السلوكية التي تهتم بدراسة وقياس ما يمكن ملاحظته ورفض غير ذلك، فقد أثرت سلباً على دراسة التفكير، مما أدى إلى اعتقاد بعض علماء النفس، باستحالة دراسة الأحداث العقلية الداخلية كعمليات للتفكير كونها لا تخضع للملاحظة المباشرة (السيد، 1995).

وعلى مدى سنوات القرن العشرين ومع ولادة علم النفس المعرفي والاهتمام بدراسة فسيولوجيا الدماغ والأعصاب، تمت إضافة الكثير إلى فهم الباحثين لعمليات التفكير والعمل على تنميته (السيد، 1995؛ عبد الله، 2003؛ Paul, Elder, & Bartell, 2004). ولقد أظهرت فترة السبعينات مؤشرات ذات أهمية في هذا الموضوع، فكان النمو الكبير في عمليات معالجة المعلومات وعلاقتها بالمعرفة من خلال دراسة الحاسوب وعلاقته بالتفكير الإنساني (السيد، 1995).

ولعل الطبيعة المركبة التي يتميّز بها التفكير الإنساني تكمن وراء تعدد مفاهيمه وأنماطه وتركز الكثير من الكتابات الحديثة على التفكير بمفهومه العام أكثر من الاهتمام بتصنيفاته، وقد يرجع ذلك إلى تداخل أنماطه والنظرة إلى عملية التفكير كعملية متصلة مستمرة (Walters, 1990).

ويشير حبيب (1995) إلى أن معظم تصنيفات أنماط التفكير كانت تميل إلى وضع التفكير الناقد مقابل التفكير الإبداعي ووضع التفكير العملي (Practical) مقابل التفكير النظري (Theoretical) ووضع التفكير التباعدي (Divergent) مقابل التفكير التقاربي (Convergent) إلا أن روجيرو (Ruggiero, cited in Salder, 1993) يميل إلى وضع جميع تصنيفات التفكير تحت مظلتي: التفكير الناقد والتفكير الإبداعي.

التفكير الناقد:

إن مصطلح التفكير الناقد واحد من أكثر المصطلحات الذي أسيء له وأسيء استخدامه، واستخدم بشكل خاطئ أيضاً، إنه ليس التفكير السلبي أو إيجاد الأخطاء الآلية لكنه يعني التفكير الذي يقوّم (Evaluate)، ويبرر ويختبر الأفكار (Grow, 1989). ويرى بروكفيلد (Brookfield, 1997, P. 17) "أن التفكير الناقد عملية واعية يقوم بها الفرد عن وعي وإدراك ولا تتم بمعزل عن البيئة المحيطة، أي انه التفكير الذي يتأثر بالسياق الاجتماعي والنظامي الذي يتم فيه". ويقول بيترس (Petress, 2004) إن التفكير الناقد يدعى في بعض الأحيان بالتفكير الموجه (Directed Thinking) لأنه يركز على النواتج المرغوبة.

ويشير إنيس (Ennis, 1993) إلى أنه بالرغم من تنصيب ديوي لموضوع التفكير الناقد على أنه هدف للتربية عبر القرن العشرين منذ عام 1910 إلا أنه لم يعمل الكثير لهذا الموضوع حتى بدايات عقد الثمانينات حيث زاد الاهتمام بالتفكير الناقد مع بعض الاهتمام بتقييمه. ولقد تمثلت زيادة الاهتمام بالتفكير الناقد كما يشير بول ونوسج (Paul & Nosich, 1992) من خلال الازدياد في عدد المؤتمرات التي عقدت في العالم حول موضوع التفكير الناقد في فترة الثمانينيات من القرن الماضي.

ويميز بول (Paul) ما بين التفكير الناقد الضعيف (Weak Sense) والتفكير الناقد القوي (Strong Sense) حيث إن الناس الذين يستخدمون مهاراتهم في التحليل والإدلاء بالحجة ليهاجموا بشكل رئيسي، وليبيّنوا عدم صحة من لا يتفقون معهم في الرأي، أناس يستخدمون التفكير الناقد الضعيف. أما الذين يستخدمون التفكير الناقد القوي فهم لا يلتزمون روتينياً بوجهة نظرهم فقط، بل إنهم يدركون ضرورة وضع فرضياتهم وأفكارهم للفحص من خلال أقوى أشكال المعارضة التي توجه ضدها (مارزانو وآخرون، 2004؛ McGuinness, 1993).

من خلال اطلاع الباحثة على الأدب التربوي المتعلق بالتفكير الناقد تم حصر 61 تعريفاً للتفكير الناقد، ولم تكن تلك التعريفات منفصلة، أو متباعدة أو مستقلة، بالرغم من اختلاف المنطلقات الفلسفية والنظرية لها، بل كان بها كثير من التداخل حيث تضمنت مجموعة كبيرة من المهارات المترابطة مع بعضها مثل التأمل والتحليل والتفسير، ومن أبرز هذه التعريفات ما جاء به ديوي (Dewey,1930, P.5) بأنه "التأمل (Reflection) الذي يتطلب النشاط، والمثابرة، والعناية والاهتمام بأي رأي أو معتقد على ضوء خلفية معينة". أما بوليت (Polette, 1982, P. 35) فيعرفه بأنه "التفكير الذي يتطلب استخدام المستويات المعرفية العليا الثلاثة في تصنيف بلوم: التحليل والتركيب والتقويم". ويعرّفه باير (Beyer, 1985, P. 272) بأنه "عملية تقرير حقيقة المعرفة، ودقتها وقيمتها، والحكم على الأخبار المستندة إلى مصادر مقبولة، وفحص المواد في ضوء الدليل ومقارنة الحوادث، ثم الاستنتاج".

ويعرفه برون وكيلي (Browne & Keeley, 1988, P. 306) بأنه "القدرة على تقويم (Evaluate) المعلومات". ويعرفه لبمان (Lipman, 1988, P. 7) بأنه "التفكير الذي يتصف بالحساسية للموقف وباشتماله على ضوابط تصحيحية وذاتية وباعتماده على محكات في الوصول إلى الأحكام". وأما ماكبيك (McPeck, 1990, 10) فيعرفه بأنه "الميل والمهارة في الانهماك بنشاط في عملية الشك التأملي". ويعرفه كل من اودال ودانيلز (Udall & Daniels, 1991, P. 5) بأنه "حل المشكلات أو التحقق من الشيء وتقييمه بالاستناد إلى معايير متفق عليها مسبقاً".

ويعرفه شافيه (Chaffee, 1992, P. 25) بأنه "الجهد الهادف المنظم والاختبار الحذر لتوضيح وتحسين فهم العالم المحيط". أما إنيس (Ennis, 1993, P. 179) فيعرفه بأنه "التفكير التأملي الذي يركز على اتخاذ القرار حول ما يعتقده الفرد أو يعمله". وعرّفه كل من وارنك

وإنش (Warnick & Inch, 1994, P. 11) بأنه "القدرة على اكتشاف المشكلة، أو السؤال، أو الموقف والعمل على تكامل كل المعلومات المتاحة، للوصول للحل وتبرير الموقف". وتم بناء اختبار كاليفورنيا بالاعتماد على تعريف أجمعت عليه مجموعة من الخبراء لمفهوم التفكير الناقد والذي ينص على أنه "عملية الحكم الهادف، المنظم ذاتياً، والمحرك المعرفي الذي يؤدي إلى حل المشكلات واتخاذ القرارات" (Facione, and Facione, 1998, P. 2)، ويعرفه هالبرن (Halpern, 2001, P. 254) بأنه "التفكير الغرضي (Purposeful) المبرر الموجه نحو الهدف الذي يتضمـــــــن حل المسألة، وصياغة الاستدلالات، وحساب الاحتمالات (Calculating Likelihoods) وعمل القرارات". وهو الوحيد الذي تفرد في مهارة حساب الاحتمالات من بين التعريفات التي تم الاطلاع عليها.

أهمية التفكير الناقد:

إن التطور التكنولوجي الهائل وتعدد الاختيارات أمام الفرد، يطرح الشعور بمسؤولية الاختيار واتخاذ القرار. والقدرة على الاختيار الجيد تتضمن بالضرورة القدرة على قياس البدائل وتقييمها تقييما صحيحاً، ولزيادة تلك القدرة يحتاج الفرد إلى مهارات التفكير الناقد لصقل الإبداع في صنع القرار وحل المشكلة والقدرة على الاختيار.

ويؤكد جلسر (Gilser, 1997, P. 87) على أن التفكير الناقد أهم مهارة تلزم مستخدم الإنترنت، لفرز المهارات الخاطئة أو غير التامة. ويؤكد وارن (Warren, cited in

(Jenkins, 1998 أنه لتحقيق احتياجات القرن الحادي والعشرين يعد التفكير الناقد من الأمور الأساسية اللازمة لمدرسة المستقبل.

ويؤكد بول (Paul, cited in Black, 2004) على أن يتعلم الطلبة كيف يفكروا بدل أن نطعمهم المعلومات بالملعقة ونسترجعها باختبارات التذكر والتي سرعان ما يتم نسيانها بعد ذلك. ولا ينكر سترنبرغ (Sternberg, 2004, P. 67) دور المعرفة في عملية التفكير الناقد، حيث يقول: "يحتاج الأطفال والراشدون على السواء قاعدة من المعرفة ليتمكنوا من التفكير الناقد السليم، لكنهم ليسوا بحاجة إلى أن يكونوا مخزناً متحركاً للمعلومات".

ويؤكد بياجيه (Piaget, cited in Wise, 2004) على أن أحد أهداف التربية هو تشكيل العقول وجعلها ناقدة. ويحذر من الشعارات، والآراء الجماعية، والأفكار الجاهزة الصنع، ويؤكد أنه على الأفراد أن يكونوا قادرين على النقد والتمييز بين الحقيقة والاعتقاد.

ويؤكد عدد من الباحثين على أن مهارات التفكير الناقد تعد عنصراً هاماً في التعليم المدرسي والجامعي ويجب أن تدرّس (Barru, Blair, Johnson, McPeck Sigel & Postmman cited in Walters, 1990)، فقد أدى الاهتمام بالتفكير الناقد في الولايات المتحدة الأمريكية إلى ظهور برامج تعليمية لتنمية التفكير الناقد لدى الأفراد (السيد، 1995؛ DeManuele, 2002; McGuinness,1993). وقامت بعض الجامعات

أيضاً بإضافة مساق لتدريس مهارات التفكير الناقد كمتطلب أساسي للجامعة (Halpern, cited in Burbach, Matkin & Fritz, 2004). ويقول جلدر (Gelder, 2005) أن مهارة التفكير الناقد صعبة، لكنها بالتأكيد ليست مستحيلة وتحتاج إلى تدريب وممارسة، وأن الممارسة تؤدي إلى الإتقان.

ويؤكد نورس (Norris, cited in Walters, 1990) على حق الطلبة في تعلم مهارات التفكير الناقد. ويضيف والترز (Walters, 1990) أن التفكير الناقد أصبح صناعة تربوية مزدهرة.

وفي الأردن أكد قانون التربية والتعليم رقم (3) على أن من أهم أهداف التربية في الأردن تنمية مهارات التفكير الناقد لدى المتعلمين (وزارة التربية والتعليم، 1994).

وتشير بعض الدراسات (السيد، 1995; Alter & Egan, 1997; Grow, 1989) إلى إخفاق الغالبية العظمى من خريجي الجامعات في التعامل بكفاءة مع المشكلات التي تتطلب مهارات التفكير المجرد، بالرغم من أن الفرد يدخل مرحلة العمليات الرسمية (Formal Operations) بدءاً من عمر 11 سنة، حيث تتميّز هذه المرحلة بقدرة الفرد على تبرير الأفكار المجردة، الافتراضية، العكسية، ويتمكن من صياغة واختبار الفرضيات المتعددة والمهتمة بربط السبب مع النتيجة، ويستطيع أيضاً في هذه المرحلة فصل وضبط المتغيـرات (Piaget, cited in Halford, 1982; Ormrod, 1995). ويوضح ديوي (Dewey, cited in Temple, 2000, P. 313) في كتاباته (مدارس الغد) (The schools of Tomorrow) سبب ذلك بقوله: "إذا قمنا بتدريب الطالب على أخذ الأوامر وتنفيذها كما قيلت له، فإننا نضع الحواجز التي لا تقهر في طريق التغلب على عيوب نظامنا الحالي". ويؤكد على تلك المشكلة أيضاً كون (Kuhn, cited in Burbach, Matkin & Fritz, 2004)

بقوله أن الغالبية العظمى من الناس لا يدعمون آراءهم بالأدلة الموثوقة أثناء مناقشاتهم وادعاءاتهم، وهذا يدلل على أهمية تعلم الإنسان مهارات التفكير الناقد. ويؤكد دي بونو (1997) على أن التفكير الناقد أداة مهمة من أدوات الاستبصار، والاستبصار هو الطريقة الفعالة في تغيير الأفكار.

ويؤكد مكبرايد (McBride, cited in Onwuegbuzie, 2001) أن الهدف الهام للتعليم العالي هو تطوير وتحسين مهارات التفكير الناقد. ولقد بينت نتائج بعض الأبحاث والدراسات أن امتلاك المعرفة بالطرق العلمية لا يضمن أن يطبق الطلبة التفكير الناقد ويمارسوه (Beck, Bennett, Mcleod, Molyneaux,Hagman, cited in Alter & Egan, 1997).

وأما فيفيانو (Viviano, cited in Temple, 2000) فيؤكد أن التفكير الناقد وليس امتلاك المعرفة هو معيار النجاح. ويرى آخرون أن التفكير الناقد متنبئ جيد للنجاح في المستقبل أو النجاح الأكاديمي في الجامعة (Williams, Oliver Allin, Winn & Booher, cited in Williams & Stockdale, 2003).

ولتأهيل وإعداد قادة عسكريين للمستقبل، يوصي الباحثون بتنمية مهارة التفكير الناقد لطلبة الكليات العسكرية لمساعدتهم على صنع القرار وحل المشكلة، والذي يعتبر ناتج مرغوب في هذا الميدان الذي يمتاز بالتعقيد والتحدي (McCarthy, 2004). ويؤكد بول (Paul, cited in Gadzella & Masten, 1998) على أهمية تعليم التفكير الناقد من أجل مجتمع حر.

مهارات التفكير الناقد:

يتشكل التفكير الناقد من مجموعة من المهارات الخاصة التي تستخدم بصورة منفردة أو مجتمعة من دون التقيد بتسلسل معين في أثنـــــاء الأداء (Grow, 1989). وقد حدد إنيس (Ennis, 1985, P. 45) 17 مهارة في التفكير الناقد هي:

1. التركيز (Focusing).

2. تحليل المناقشات والحجج (Analyzing arguments).

3. طرح الأسئلة الجيدة (Asking good questions)

4. تعريف المصطلحات والحكم على التعريفات (Defining terms & judging the definitions).

5. التعرف إلى الافتراضات (Identifying assumptions).

6. الحكم على مصداقية مصادر المعلومات (Judging the credibility of source)

7. ملاحظة وتقييم تقارير الملاحظات(Observing and judging observation reports).

8. الاستنتاج (Deduction).

9. يستقري ويحكم على النتيجة المستخلصة من عملية الاستقراء (Inducing and judging inductions).

10. الحكم القيمي (Making and judging value judgments).

11. الاستدلال (Inference)

12. التكامل (Integration).

13. الترتيب (Orderly manner).

14. الحساسية تجاه الآخرين (Sensitivity to others).

15. البيان والبلاغة (Rhetorical).

16. المصطلحات الخاطئة (Fallacy" Labels").

17. التفاعل مع الآخرين (Interacting with others).

أما باير (Beyer, 1985, P. 272) فقد حدد عشر مهارات رئيسية للتفكير الناقد وهذه المهارات هي:

1. التمييز بين الحقائق التي يمكن إثباتها أو التحقق من صحتها وبين الادعاءات أو المزاعم الذاتية.
2. التمييز بين المعلومات والادعاءات والأسباب ذات العلاقة بالموضوع وتلك التي تقحم على الموضوع ولا ترتبط به.
3. تحديد مصداقية مصدر المعلومات.
4. تحديد الدقة الحقيقية للخبر أو الرواية.
5. التعرف على الادعاءات أو البراهين والحجج الغامضة.
6. التعرف على الافتراضات غير الظاهرة أو المتضمنة في النص.
7. تحري التحيز.
8. التعرف على المغالطات المنطقية.
9. التعرف على أوجه التناقض أو عدم الاتساق في مسار عملية الاستدلال من المقدمات إلى الوقائع.
10. تحديد درجة قوة البرهان أو الادعاء.

وقسم كل من سكريفن وبول (Scriven & Paul, 1992) التفكير الناقد إلى (35) مهارة ضمن ثلاثة أبعاد هي:

أ. البعد الانفعالي (Affective):

English	العربية
1.Thinking Independently	1. التفكير باستقلالية
2.Developing insight into egocentricity or socioecentricity	2. تطوّر التبصر الفردي التمركز أو المجتمعي التمركز
3.Exercising fair-mindedness	3. ممارسة العقلية العادلة
4.Exploring thoughts underlying feeling & feeling underlying thoughts	4. اكتشاف الأفكار المتضمنة في المشاعر والمشاعر المتضمنة في الأفكار
5.Developing intellectual humility & Suspending Judgment	5. تطوير التواضع الفكري وتجنب الأحكام
6.Developing intellectual courage	6. تطوير الشجاعة الفكرية
7.Developing intellectual good faith or integrity	7. تطوير تفكير النوايا الحسنة
8.Developing intellectual perseverance	8. تطوير المثابرة الفكرية
9.Developing confidence in reason	9. تطوير الثقة في السبب

ب. البعد المعرفي (Cognitive (Macro- Abilities:

English	العربية
10.Refining generalization & avoiding over simplifications	10. تهذيب التعاميم وتجنب السطحية
11.Comparing and analogous situations: Transferring insights to new contexts	11. مقارنة المواقف المتشابهة: نقل التبصر إلى سياقات جديدة
12.Developing one's perspective: creating or exploring believes, arguments or theories	12. تطوير وجهة نظر الشخص: خلق أو استكشاف الاعتقادات أو الحجج أو النظريات
13.Clarifying issues, conclusions or believes	13. توضيح القضايا والنتائج أو الاعتقادات
14.Clarifying & analyzing the meaning of the words or phrases	14. توضيح وتحليل معاني الكلمات أو الجمل
15.Developing criteria for evaluation: clarifying values & standards	15. تطوير معايير التقويم: توضيح القيم والمعايير
16.Evaluating the credibility of sources of information	16. تقويم مصداقية مصادر المعلومات
17.Questioning deeply: raising & pursuing root or significant questions	17. التساؤل بعمق: طرح ومتابعة الأسئلة الأساسية الهامة
18.Analyzing or evaluating arguments	18. تحليل أو تقويم الحجج
19.Generating or assessing solutions	19. توليد أو تقييم الحلول
20.Analyzing or evaluating actions or policies	20. تحليل أو تقويم الأفعال أو السياسات
21.Reading critically: Clarifying or critiquing texts	21. القراءة الناقدة: توضيح أو نقد النصوص
22.Listening critically: the art of silent dialogue	22. الاستماع الناقد: فن الحوار الصامت (الداخلي)
23.Making interdisciplinary connections	23. عمل روابط متعددة الميادين المدنية
24.Practicing Socratic discussion	24. ممارسة المناقشة السقراطية
25.Reasoning dialogically: comparing perspectives, interpretations, or theories	25. حوار عقلاني أو مبرر: مقارنة وجهات النظر، التفسيرات أو النظريات
26.Reasoning dialectically	26. التبرير الجدلي

ج. البعد المعرفي (Micro skills) Cognitive:

27. مقارنة المثاليات مع الواقع	27.Comparing & contrasting ideals with actual practice
28. التفكير بدقة عن التفكير: استخدام مفردات دقيقة	28.Thinking precisely about thinking: using critical vocabulary
29. ملاحظة التشابهات والاختلافات المهمة	29.Noting significant similarities & differences
30. فحص أو تقويم الافتراضات	30.Examining or evaluating assumptions
31. تمييز الحقائق ذات الصلة من غير ذات الصلة	31.Distinguishing relevant from irrelevant facts
32. عمل استدلالات وتنبؤات أو تفسيرات معقولة	32.Making plausible inferences, predictions or interpretations
33. إبداء الأسباب وتقويم الدليل والحقائق المزعومة	33.Giving reasons & evaluating evidence & alleged facts
34. التعرّف على التناقضات	34.Recognizing contradictions
35. اكتشاف التضمينات والعواقب	35.Exploring implications & consequences

وبالإضافة للمهارات السابقة فقد أورد كثير من الباحثين مهارات أخرى منها:

1. مهارة جمع المعلومات وتنظيمها (McWhorter & Collins, 1992).
2. المعالجة الناقدة للمعلومات (Kerka, 1992 ; McWhorter & Collins 1992).
3. تشكيل المفاهيم (Marzano et al; cited in Sormunen & Chalupa, 1994)
4. توليد البدائل (Halpern, 2001).
5. التفسير (Facione & Facione, 1998 ;Grow, 1989; Kerka, 1992; McWhorter & Collins, 1992).
   - التصنيف (Facione and Facione, 1998).
   - استخلاص المغزى أو التضمين (Facione and Facione, 1998).
   - توضيح المعاني (Facione and Facione, 1998).
   - الحكم على منطقية الاستنتاجات في ضوء الأدلة (Watson & Glaser, 1964).
6. التلخيص (Marzano et al; cited in Sormunen & Chalupa, 1994).
7. تحديد العنوان (Brown and Salisch, 1996).

8. التوسع (Marzano et al; cited in Sormunen & Chalupa, 1994).

9. التمييز بين أنواع العلاقات الآتية:

- السبب والنتيجة.
- الارتباطية.
- علاقة التناظر (Analogies) (Potts, 1994).

10. التقويم (Evaluation) (Facione & Facione, 1998; Miller, cited in Kerka, 1992).

- النقد (King, cited in McWhorter & Collins ,1992):
- تحري التحيز.
- تقويم الادعاءات (Facione & Facione, 1998; McWhorter & Collins, 1992).
- تقويم مختلف وجهات النظر (Potts, 1994).
- تقويم المناقشات والحجج (Facione & Facione, 1998).
- إيجاد وتقويم الحلول أو الطرق البديلة للتعامل مع المشكلة (Watson &Glaser, 1964).
- الإثبات (Burbules, cited in Siegel, 1999).
- البحث عن الدليل الذي يميز الاختيار الأفضل.
- وزن كفاية البيانات أو الدليل.

11. اتخاذ القرار (Kerka, 1992 ; Sawin, 2004):

- مهارة صنع المعايير لإصدار القرار (Sawin, 2004).
- العدالة أو الإنصاف الفكري (Celuch, and Slama, 1999).

12. التبرير الإحصائي والاحتمالي (Halpern, 2001; Sawin, 2004).

- استخلاص الفكرة أو المغزى من الرسم والجدول والبيانات.
- تفسير الرسم والجدول والبيانات.
- التنبؤ من الرسم والجدول والبيانات وعمل القرارات.
- القدرة على التقدير.
- تجنب التبرير أو التفكير أو الاستنتاج والتفسير الخاطئ من البيانات.
- الحكم على جودة الدليل الإحصائي.
- التبرير الاحتمالي (Royalty, 1995).
- التبرير القياسي(Royalty, 1995).

13. التنظيم الذاتي: (Facione & Facione, 1998):

- الفحص الذاتي (Facione & Facione, 1998).
- التصحيح الذاتي (Facione & Facione, 1998).
- إدارة الذات (Prosser et al; cited in Walkner & Finney, 1999).
- مراقبة الذات (Self-Monitoring) (Brown & Salisch, 1996).
- القدرة على الاعتراف بالخطأ (Van Duzen, cited in Irving & Williams, 1995).
- يعرف ماذا يعمل (Arrgyris, cited in Irving & Williams, 1995.).
- الانضباط الذاتي (Royalty, 1995)
- موجه ذاتياً (Brookfield cited in McGuinness, 1993).
- ضبط داخلي (Internal Control) (Jenkins, 1998).

- تقبل الحشد (الكل) (Tolerance to follow the crowed) (Vincenti, cited in Harrigan & Vincenti, 2004).
- الفضول الفكري (Intellectual curiosity) (Vincenti, cited in Harrigan & Vincenti, 2004).
- المرونة التكيفية (القدرة على التكيف مع تغير الظروف) (Creter & Oaksford, cited in Astleitner, 2002).
- التحرر من الجمود.
- إعادة تفسير المعلومات.
- التفكير التكيفي (Creter & Oaksford, cited in Astleitner, 2002).
- التماسك أو الترابط المنطقي (Coherence) (The Role of Critical Thinking, 1997).

14. التركيز: (Ennis, 1985)

- تحديد التوجهات والأهداف (Jenkins, 1998).
- النظر إلى الأشياء بتوسع أكبر (Walkner & Finney, 1999).
- النظر إلى الأشياء بعمق أكبر (Walkner & Finney, 1999).

15. القدرة على كشــف الخدع وردعه (Fraud deterrence & detection ability) (Jenkins, 1998).

16. تكييف (توجيه) المسألة (Problem orientation) (Lee, cited in Kerka, 1992)

17. النزعة إلى الشك (Skepticism) (Royalty, 1995).

18. الإقناع (Persuasion) (Sormunen & Chalupa, 1994).

19. المعاني، أو الدلالة اللفظية (general semantic) (Sawin ,2004).

20. الاستدلال (Inference) (Facione & Facione, 1998; Salder, 1993).

- تقصي الأدلة.
- تخمين البدائل.
- التوصل إلى الاستنتاجات.
- الحكم على صحة الاستنتاج.
- الاستدلال الاستنتاجي (Deductive Inference) (Rolyalty, 1995).

21. الاتصال والتواصل (Kerka, 1992; Brown & Salisch, 1996).

- الثقة الفكرية (Celuch, and Slama, 1999).
- التعلم مدى الحياة (Celuch, and Slama, 1999).
- التعامل بفعالية مع تغيرات العالم السريعة (Celuch, and Slama, 1999).
- المثابرة (Perseverance) (Katsioloudes & Tischio, 2001)
- الاعتناق والتقمص العاطفي (Celuch, and Slama, 1999).
- التعلم كيف تتعلم (Kerka, 1992).
- عدم اللجوء إلى التعميمات الزائدة (المفرطة) (The Role of Critical Thinking, 1997).
- عدم اللجوء إلى عامل الحشد (المجموع) (The Role of Critical Thinking, 1997).
- القدرة على عمل التسويات (The Role of Critical Thinking, 1997)

22. التأمل (Reflection) (Goyne, 2001).

23. العقل المتفتح (Open – Mindedness) (Temple, 2000).

- التفكير غير المتحيز.
- رفض الانغلاق المبكر.
- معايير عقلانية (تكوين واستخدام معايير عقلانية) (Lee, cited in Kerka, 1992; Paul, 1998).
- النظر إلى الأشياء من وجهات نظر مختلفة (Bowden & Mazten, cited in Walkner & Finney, 1999).

24. حسن التبصر (kerka, 1992).

25. تطوير طرق جديدة لرؤية المواقف (الظواهر) (Kerka, 1992).

26. القدرة على تقبّل الغموض أو الالتباس (Miller, cited in Kerka, 1992).

27. التحليل (Grow, 1989; Kerka, 1992).

- تجزئة المشكلة (Grow, 1989).
- التعرف على الأنماط (Pattern Recognition).
- فحص الآراء (Facione & Facione, 1998; McWhorter & Collins, 1992).
- اكتشاف وتحليل الحجج (Facione & Facione, 1998).
- التعمق (Pursuing Roots) (Vincenti, cited in Harrigan & Vincenti, 2004).
- القدرة على التعريف بنتائج الحجج والمناقشات والتمييز بين النتائج الإيجابية والسلبية منها (The Role of Critical Thinking, 1997)

- التعريف بالمناقشة أو الحجج الأخلاقية وأن يميز بين الحجة الأخلاقية وغير الأخلاقية (The Role of Critical Thinking, 1997).

28. نقل المعرفة من العالم الأكاديمي إلى عالم العمـــل (Brans, Ford & Stein, cited in Sormunen & Chalupa, 1994).

29. التركيب (McWhorter & Collins, 1992).

30. العدالة (Fairness) (Royalty, 1995).

31. مزود بالخبرة ((Jenkins, 1998.

32. نقل المعلومات (المعرفة) (Lee, cited in Kerka, 1992).

33. الحكم الحدسي (Intuitive Judgment) (Crater& Oaksford, cited in Astleitner, 2002).

34. التفكير المفتوح (Open Thinking) (Temple, 2000).

35. النمذجة (Modling) (Celuch and Slama, 1999).

36. الشرح (Explanation) (Facione & Facione, 1998).

- وضع النتائج.
- تبرير الإجراءات.
- تقديم الحجج.

معايير التفكير الناقد:

هناك مجموعة من المعايير التي تستخدم للحكم على جودة التفكير الذي يستخدمه الفرد في معالجة مشكلة ما، ومن أبرز تلك المعايير ما ورد عن ألدر وبول (Elder&Paul, 2001).

1. الوضوح (Clarity)

- هل تستطيع أن توضح أكثر.
- هل تقدر أن تعطي مثال.

2. الصحة (Accuracy).

- كيف يمكن أن تفحص ذلك.
- كيف يمكن التأكد من صحة ذلك.

3. الدقة (Precision)

- هل يمكن أن تكون أكثر تحديداً (Specific)
- هل يمكن أن تعطي تفصيلاً أكثر.

4. وثيق الصلة بالموضوع (Relevance)

- كيف يرتبط هذا مع المشكلة.
- كيف يساعدنا ذلك في معالجة المشكلة.

5. العمق (Depth)

- ما العوامل التي تجعل هذه المشكلة صعبة.

6. الاتساع (Breadth)

- هل تحتاج لأخذ وجهة نظر الآخرين.
- هل هناك أسلوب آخر للتعامل مع المشكلة.

7. المنطق (logic)

- هل تتلاءم الفقرة الأولى لديك مع الأخيرة.
- هل ما تقوله يحمل دليلاً.

8. الدلالة أو الأهمية (Significance)

- هل هذه المشكلة هي الأكثر أهمية لأخذها.
- هل هذه الفكرة الرئيسة للتركيز عليها.
- أي من هذه الحقائق facts أكثر أهمية.

9. العدالة أو عدم التحيّز (Fairness).
10. الاتساق أو الثبات (Consistency).

خصائص المفكر الناقد:

احتلت سمات وخصائص المفكرين الناقدين اهتمام كبير من قبل العديد من البــاحثين والعلماء، ولقد حدد فرت (Ferrett, cited in Petress, 2004) صفات المفكر الناقــد كمــا يأتي:

1. يسأل أسئلة ذات صلة.
2. يقيم العبارات والحجج.
3. قادر على التعرّف على النقص في المعلومات.
4. يمتلك إحساساً بالفضول.
5. يهتم بإيجاد حلول جديدة.
6. قادر على التعريف بمعايير واضحة لتحليل الأفكار.
7. قادر على فحص الآراء والاقتراحات، والمعتقدات، ومقارنة ذلك مع الحقائق.
8. يستمع باهتمام للآخرين، وقادر على إعطاء تغذية راجعة.
9. يرجئ إصدار الأحكام لحين تجمّع كل الحقائق لديه.
10. يبحث عن الدليل الذي يدعم افتراضه واعتقاده.
11. قادر على تعديل آرائه عند إيجاد حقائق جديدة.

12. يبحث عن البرهان.

13. يختبر المشاكل بدقة.

14. قادر على رفض المعلومات غير الصحيحة أو غير ذات الصلة.

15. يقيم ذاته باستمرار.

وأكد إنيس (Ennis, 1993) على بعض الخصائص التي ذكرها فرت (Ferrett)، وأضاف صفات أخرى للمفكر الناقد هي:

1. الحكم على مصداقية المصادر.
2. الحكم على جودة الحجة بما فيها مصداقية أسبابها وافتراضاتها ودليلها.
3. تطوير أو دحض موقف في قضية.
4. يسأل أسئلة واضحة ومناسبة.
5. يخطط لتجارب ويحكم على التصاميم التجريبية.
6. يعرف المصطلحات بطريقة ملائمة للنص.
7. يحاول أن يكون حسن الاطلاع (Well informed.).
8. متفتح العقل.
9. حذر في استنتاجاته.

حركة قياس التفكير الناقد:

إن التفكير الناقد مهارة قابلة للتحسن بالتدريب والتعليم (Celuch, & Slama, 1999) وبناء على ذلك فإنه لا بد من مقياس يحدد مقدار التقدم في التفكير الناقد لدى الشخص بعد تدريبه وتعليمه استراتيجيات تحسين مهارة التفكير الناقد. ويتفق كل من سكريفن وبول (Scriven & Paul, 1992) مع بروكفيلد (Brookfield, 1997) على أن مهارة التفكير

الناقد ليست عامة وشاملة في أي شخص بل تعتمد في درجة وجودها على نوعية وعمق الخبرة، مما يستوجب قياسها والعمل على تطويرها.

ولقد أكد ديوي (المشار إليه في مارزانو وآخرون، 2004) أن الطريق الوحيد المباشر لتحقيق تحسن مستمر في طرق التعلم والتعليم يتمركز في الظروف التي تؤدي إلى محاكمة التفكير وتعزيزه وقياسه. وقديماً كانت أداة قياس التفكير هي الاستبطان (Introspection) حيث كان تناول موضوع التفكير يتسم بالعرض الشمولي لجميع أنواع التفكير من دون التركيز على أي نوع (السيد، 1995).

إن قياس التفكير الناقد ارتبط ارتباطاً وثيقاً ببرامج تنميته، فالكثير من المحاولات التي بذلت لتحليل مكونات التفكير الناقد، وتحديد المهارات المكوّنة له، قد أعانت على تسهيل عملية قياسه كما أعانت في الوقت نفسه على إمكانية وضع البرامج المختلفة لتنمية هذه المهارات أو المكونات التي اشتمل عليها.

وقد أسفرت محاولات الباحثين لبناء أدوات لقياس التفكير الناقد عن عدد من اختبارات التفكير الناقد، ومن أكثر هذه الاختبارات شيوعاً، مرتبة وفق التسلسل الزمني لها ما يأتي:

1. **اختبار واتسون – جليسر (Watson – Glaser Test)**: أعد هذا الاختبار واتسون وجليسر سنة 1941. وله نموذجان متكافئان يتكون كل منهما من خمسة اختبارات فرعية تقيس تمييز المسلمات، والاستنتاج، والاستقراء، والتفسير، وتقويم الحجج. ويمكن استخدامه بدءاً من الصف التاسع. تدور فقراته حول مشكلات اجتماعية محيطة بالأفراد (Watson & Glaser, 1964).

2. **اختبار التبرير المنطقي (Logical Reasoning Test)**: أعد هذا الاختبار هيرتزكا وجليفورد (Hertzka & Guilford) سنة 1955. ويستهدف الأفراد بدءاً بالمرحلة الثانوية. ويتكون من نموذجين متوازيين تمت معايرتهما لمجتمعين مختلفين من طلبة المرحلة الثانوية العليا، والكليات (Paul & Nosich, 1992; Sormunen & Chalupa, 1994).

3. **اختبار كورنل للتبرير (النموذج X)(Cornell Class Reasoning Test, Form X)**: أعد هذا الاختبار إنيس وجاردينر ومورو وبولص وريبنج (Ennis, Gardiner, Morrow, Paulus & Ringe) سنة 1964. ويقوم على اختبار قدرة الفرد على الاستنتاج. ويستهدف الطلبة من الصف الرابع حتى الكلية (Paul & Nosich; 1992; Sormunen & Chalupa, 1994).

4. **اختبار كورنل للتبرير الشرطي (النموذج X) (Cornell Conditional Reasoning Test, Form X)**: أعد هذا الاختبار إنس وجاردينر وجازيتا ومورو وبولص وريبنج (Ennis, Gardiner, Guzzetta, Morrow, Paulus & Ringe) سنة 1964. يستهدف الطلبة من الصف الرابع حتى الكلية (Paul & Nosich, 1992; Sormunen & Chalupa, 1994).

5. **اختبار الحكم: الاستدلال المنطقي وإدراك الافتراضات: (Judgment: Deductive Logic & Assumption Recognition Test)** أعد هذا الاختبار شافير وستيجر (Shaffer & Steiger) سنة 1971، ويستهدف الطلبة من الصف السابع وحتى الثاني عشر، وقد بني على أساس محكي المرجع (Criterion - Reference) (Paul & Nosich, 1992; Sormunen & Chalupa, 1994).

6. اختبار كري للتفكير الناقد (Curry Test of Critical Thinking): أعد هذا الاختبار كري (Curry) سنة (1971) ويستهدف الفئة من الصف التاسع ولغاية الصف الثاني عشر ويقوم على اختبار القدرات الآتية:

- الحقائق والآراء.
- الخطأ والصح.
- تكوين الافتراضات.
- التعرف إلى البيانات الناقصة.
- القياس أو التناظر غير الحقيقي (Landis & Michael, 1981).

7. اختبار روس للعمليات المعرفية العليا (Ross Test of Higher Cognitive Processes): أعد هذا الاختبار روس (Ross) سنة 1976، ويستهدف الطلبة من الصف الرابع وحتى السادس، ويقوم على اختبار القدرة على القياس اللفظي، والاستنتاج، والتعريف بالافتراضات، والتفسير (Paul & Nosich, 1992).

8. اختبار مهارات الاستقصاء (Test of Inquiry Skills): أعد هذا الاختبار فريسر (Fraser) سنة 1979، ويستهدف الطلبة من الصف السابع وحتى الصف العاشر، ويقوم على اختبار القدرة على استخدام المراجع والفهارس في المكتبة، والقدرة على التفسير، ومعالجة المعلومات، والمقاييس، والرسومات والصور (Paul & Nosich, 1992).

9. اختبار تقييم الملاحظات (Test on Appraising Observations): أعد هذا الاختبار نورس وكينج (Norris & King) سنة 1983، ويستهدف الطلبة من الصف السابع حتى الكلية. ويقوم على إعطاء مجموعة من المبادئ (Principles) حيث يتم اختيار الفقرات بعد إجراء مقابلات مع الطلبة (Paul & Nosich, 1992; Sormunen & Chalupa, 1994).

10. **اختبار نيوجرسي لمهارة التبرير (New Jersey Test for Reasoning Skill):**

أعد هذا الاختبار شبمان (Shipman) سنة 1983، ويستهدف الطلبة من الصف الخامس حتى المستوى الجامعي، إذ أن نصف فقرات الاختبار معنية بالاستنتاج (deduction)، أي يمثل وجهاً واحداً من أوجه التفكير الناقد (Paul & Nosich, 1992; Sormunen & Chalupa, 1994).

11. **اختبار إنيس – وير المقالي للتفكير الناقد (Ennis – Weir Critical Thinking essay test):** أعد هذا الاختبار إنيسووير (Ennis & Weir) سنة 1985 وهو اختبار مقالي، متعدد الأوجه (Multi – aspect test) صمم لطلبة المرحلة الثانوية والكليات. ويصحح الاختبار يدوياً ويعد أداة تشخيصية وبحثية، اعتمد هذا الاختبار في قياس التفكير الناقد على إعطاء المزيد من الحرية لتقويم المناقشات وتمحيص العبارات، إذ على المستجيب أن يكوّن الاستجابة بنفسه (Paul & Nosich, 1992; Sormunen & Chalupa, 1994).

12. **اختبار كورنل للتفكير الناقد (Cornell Critical Thinking Test):** أعد هذا الاختبار إنيس وميلمان (Ennis & Millman) سنة 1985 وله مستويان: المستوى X للطلبة من الصف الرابع حتى الكلية، والمستوى Z مناسب لطلبة المرحلة الثانوية العالية وطلبة الكليات والراشدين. يركز كلا المستويين بشكل أساسي على أوجه التقويم الخاصة بالتفكير الناقد حيث صيغت مواقف هذا الاختبار على هيئة مناقشة حول قضية ما، تنتهي المناقشة بنتيجة، يطلب من الفرد المستجيب أن يحكم على مدى صحة هذه النتيجة واتساقها مع الوقائع، ويقيس هذا الاختبار القدرات الآتية: الاستقراء، والمعنى، والاستنتاج، وتحديد التعريف، وتحديد المسلمات، ومصداقية العبارات. يتكون المستوى

X من إحدى وسبعين عبارة مرتبطة بقصة تدور حول مجموعة من مكتشفي الفضاء هبطت على كوكب لمعرفة أخبار مجموعة قد سبقتهم وفُقدت، ويضع أفراد الجماعة المكتشفون ملاحظاتهم وفروضهم، ويستمعون إلى تقارير رؤسائهم، ويفسرون الأحداث، على المستجيب لهذا الاختبار أن يجيب عن أسئلة اختيار من متعدد وذلك لقياس مدى ثبات المعلومات، وكفاءة المعلومة لاختبار الفروض، ومدى اتساق العبارات مع الوقائع، ثم تحديد المسلمات وقد تراوح معامل الثبات المحسوب بطريقة التجزئة النصفية لهذا الاختبار ما بين (0.69إلى 0.76) حيث ارتبط المعامل الأعلى بمستوى العمر الأعلى للعينة (السيد، 1995(Paul & Nosich, 1992; Sormunen & Chalupa; 1994.

13. **اختبار قدرة الاستدلال في الاستيعاب القرائي (Test of Inference Ability in Reading Comprehension):** أعد هذا الاختبار فيلبس وباترسون (Philips & Paterson) سنة 1987، ويستهدف الطلبة من الصف السادس وحتى الثامن، ويقوم على اختبار القدرة على الاستدلال والتفسير(Paul & Nosich, 1992).

14. **اختبار كاليفورنيا لمهارات التفكير الناقد (California Critical Thinking Skill Test):** أعد هذا الاختبار فاسيون (Facione) سنة 1990 لقياس مهارات التفكير الناقد الآتية: التحليل، والتقويم، والاستدلال، والاستنتاج، والاستقراء، ويستهدف هذا الاختبار الطلبة من الصف العاشر وحتى الكلية، ويتكون من نموذجين متوازيين يحوي كل منها (34) فقرة من نوع الاختيار من متعدد لكل فقرة أربعة بدائل (Facione, & Facione, 1998). وهناك نموذجان آخران للاختبار هما نموذج عام (1992) ونموذج عام (2000).

15. اختبار مينـــسوتا للتفكيـــر الناقـــد (Minnesota Test of Critical Thinking (MTCT)): يقوم هذا الاختبار على قياس مهارات التفكير الناقـــد الآتيـــة: التحليـــل والتقويم والاستدلال والتفسير والشرح، معتمداً بذلك على تصنيف مهارات التفكير الناقد المشتقة من التقويم والتعليم التربوي (Educational Assessment & Instruction) ويعرّف التفكير الناقد على أنه الحكم الهادف والمنظم ذاتياً (Eldman, 2002).

يشير الأدب حول هذه الاختبارات إلى أنها لا تعكس أو تقيس عمليات التفكيـــر التـــي يستخدمها المفحوص للوصول إلى الإجابة التي تمثل نواتج التفكير. وغالباً ما يكـــون البنـــاء العاملي لها مختلفا عن البنـــاء النظـــري (Millman, Tomb, cited in ،Norris, Ennis Sormunen & Chalupa, 1994) بالإضافة إلى انخفاض في مقدار الثبات لـــبعض هـــذه المقاييس وإلى الارتباطات الضعيفة بين الفقرة والمقياس (Stein, Haynes & Unterstein, 2003).

ويؤكد فرسبي (Frisby, 1991) أن على مطوري اختبارات التفكير الناقد أن يعملوا جاهدين للإجابة عن السؤال الصعب "ماذا تعني العلامة على هذا الاختبـــار"؟. وأن يركـــزوا اهتمامهم على العلاقة بين السمة موضع الاهتمام وعلامات الاختبار نفسه، وذلك في دراســـة قام خلالها بمقارنة علامات افتراضية لاختبار كورنل للتفكير الناقد للفئة المتدنية من فئة العمر 13 سنة مع علامات التفكير الناقد للفئة العليا من فئة العمر 9 سنوات، حيث وجد أن علامات الفئة الأولى (13 سنة، مستوى متدني) كانت أعلى من الفئة الثانية (9 سنوات، مستوى عالٍ) مما جعل قبول هذه النتيجة صعب لبعض المنظرين، ويبدو أن هناك ســـبباً جعـــل العلامـــة الملاحظة تزداد عبر العمر الأعلى، وربما يعزى ذلك إلى أن الاختبار يقيس شيئاً مختلفاً عن قدرات التفكير الناقد، ويؤكد الباحث بعد الدراسة المسحية التي أجراها لمجموعة من الدراسات

التي تناولت ارتباط مهارة التفكير الناقد مع العمر، أنه باستثناء الخطأ العيني، قد تكون تلك الزيادة تعكس زيادة حقيقية في السمة تجعل مهمة التفكير الناقد أسهل، أو قد تفسر تلك الفروق بشكل جزئي، حسب تصنيف بياجيه بالانتقال إلى مرحلة العمليات الرسمية (Formal Operations) أو بالزيادة التطورية في قدرة الاستيعاب القرائي، أو أن تلك الاختبارات، ربما تقيس خليطاً من سمات مختلفة عبر مراحل النمو المختلفة. وأشار فرسبي إلى أن استخدام منحنى خصائص الفقرة ربما يساعد في تحديد فقرات اختبار كورنل النموذج X التي تعكس خصائص غير مرغوبة مرتبطة مع الازدياد في مستوى الصف.

ويقول إنيس (Ennis, 1993) أنه لا يوجد أي اختبار من اختبارات التفكير الناقد المنشورة، مناسباً لكل الأغراض، ويضيف قائلاً أنه على الباحثين إيلاء تعريف التفكير الناقد أهمية أكبر، لأن التعريف يساعد الباحث في توضيح ماذا يريد أن يقيّم. ويقول أيضاً إن واحداً من الانتقادات الرئيسة لأغلب اختبارات التفكير الناقد الموجودة هي الافتقار إلى الشمول في تغطية مهارات التفكير الناقد بشكل كافي، فعلى سبيل المثال وبشكل خاص تفشل الاختبارات في اختبار أشياء هامة مثل العقل المتفتح ومصداقية المصادر. ويؤكد هذا كل من سبايسر وهانكس (Spicer & Hanks, 1995) حيث أشارا إلى النقص في توفر اختبار للتفكير الناقد يغطي المهارات بشكل كاف. ويلفت بايك (Pike, 1997) النظر إلى أن رؤية التفكير الناقد كمجموعة منفصلة من معرفة محددة المجال (Domain- Specific Knowledge) قد يكون السبب الرئيس في عدم حساسية المقاييس الحالية لآثار التعليم، والتدريب على البرامج الخاصة في تنمية مهارة التفكير الناقد.

وبسبب تطوير تعريف التفكير الناقد بشكل شامل فقط خلال السنوات العشر الأخيرة يؤكد كثير من الباحثين على أنه لم يحن الوقت بعد لتطوير أدوات تقييم جديدة تقوم فقراتها على فحص مجال واسع من القدرات (Large - Scale Abilities)، وتعالج أمثلة حياتية ومسائل وقضايا تشغل تفكير الناس بالإضافة إلى التركيز على مناسبة الاختبار لكافة مستويات التعليم ولكافة مستويات الطلبة عبر الثقافات المختلفة (Paul & Nosich, 1992).

ويضيف سكنر (Skinner, cited in Wood & Conner, 1999) أن الاتفاق حول العناصر التي يجب أن تدخل في عملية التفكير الناقد والعلاقات بين تلك العناصر ما زالت غير واضحة.

ويضيف كريبر (Kreber, 1998) أن محتوى اختبار واتسون- جليسر محط تساؤل، وغير واضح فيما إذا كان يعرض مدى كافٍ من مهارات التفكير الناقد المطلوبة للمواقف الحياتية الحقيقية.

وقد حدد كل من بول ونوسج (Paul & Nosich, 1992) معايير للحكم على قوة وصلاحية أي اختبار لقياس مهارات التفكير من الرتبة العليا وهذه المعايير هي:

1. تقييم مهارات وقدرات الطلبة في التحليل والتركيب، والتطبيق وتقويم المعلومات.
2. تركز على مهارات التفكير والتي يمكن أن تستخدم بأقصى مرونة، وعلى شريحة واسعة من الأفراد والمواقف والبيئات ومستويات التعليم.
3. تعمل على تفسير أو شرح الفروق الهامة بين الأفراد من مهارات، وعمليات، ومجالات انفعالية والتي هي أساسية لجميع الأفراد.
4. تقود إلى تحسين التعليم.

5. تعمل على توضيح الاتصال الداخلي للمعرفة والقدرة وتكون حساسة للتكامل بين المبادئ المتداخلة.

6. تعمل على تقييم مهارات متنوعة وضرورية ليكون الفرد عضواً مسؤولاً قادراً على صنع القرار في مكان عمله.

7. تعتمد على مفاهيم واضحة، وتمتلك أهدافاً ومعايير واضحة وجيدة.

8. تعمل على تكامل مهارات الاتصال وحل المسألة والتفكير الناقد.

9. تحترم تنوع الثقافات وذلك عن طريق التركيز على المهارات الأساسية المركزية العامـــة Common- Core Skills.

10. تعمل على اختبار التفكير الذي إذا اتحد مع التعلّم يعمل على تقويـــة وتعزيـــز الانهمـــاك النشط للطلبة في تكوين معرفتهم وفهمهم الخاص.

11. تركز على تقييم البنية المعرفية الأساسية للاتصال

12. تتجنب أي نقص يسمح بتعدد الأوجه أو تعقيد الموقف.

13. تمكن المربين من رؤية أنواع المهارات الأساسية اللازمة للمستقبل.

14. تتضمن فقرات تقيّم مهارات التفكير التي تختار أكثر الإجابات معقولية.

15. تتضمن فقرات تحتوي على مواقف حياتية ومشاكل وقضايا تشغل تفكير الناس.

16. أن تكون سهل الوصول إليها (Affordable).

17. تمكن المدارس والمربين من تقييم ما تم اكتسابه في أثناء تعليمهم مهـــارات التفكيـــر العليا.

كما أن معظم المقاييس الحالية- إن لم تكن جميعها-قد تـــم بناؤهـــا وفـــق النظريـــة الكلاسيكية للقياس وبذلك فإن خصائصها السيكومترية مثـــل معـــاملات الـــصعوبة والتمييـــز

والصدق والثبات، تتأثر بخصائص عينة الأفراد الذين تطبق عليهم، ولهذا تدعو الحاجة إلى بناء مقاييس جديدة تتفق والقياس الموضوعي وتتمتع بفقرات ذات خصائص سيكومترية متحررة عن الأفراد، وبتقدير قدرات الأفراد بصورة موضوعية متحررة عن الفقرات التي يستجيبون لها (Frisby, 1991). ويمكن تحقيق ذلك باستخدام نماذج نظرية السمات الكامنة (Latent Trait Theory) حيث تتميّز هذه النماذج بما يأتي:

- وجود مجموعة كبيرة من فقرات الاختبار تقيس نفس السمة ويكون تقدير قدرة الفرد مستقلاً عن عينة الفقرات التي تطبق عليه.
- وجود مجتمع كبير من الأفراد، تكون الخصائص السيكومترية للفقرات (مثل معاملات الصعوبة والتمييز) مستقلة عن عينة الأفراد التي استخدمت في تقدير هذه الخصائص.
- وجود معامل إحصائي يبيّن مدى دقة تقدير قدرة كل فرد بواسطة فقرات الاختبار، وربما يختلف هذا المعامل من فرد إلى آخر (Hambleton & Swaminathan, 1985).

وحيث إن توزيع علامات الفقرات الخاص بمستوى قدرة معين يجب ألا يختلف من مجتمع لآخر، مما يعني أن منحنى خاصية الفقرة لا يتغير (Invariant) بتغير مجتمعات الأفراد التي استخدمت في تعريف وتحديد الفضاء الكامن، ويعد عدم تغير منحنى خاصية الفقرة، وما يتبع ذلك من عدم تغيير للمعالم التي تحددها إحدى الميزات التي تتميّز بها نماذج السمات الكامنة، حيث تتسم هذه النماذج بأنها نماذج ذات فقرات متحررة من أثر قدرات الأفراد (Person-Free Item)، وأن قدرات الأفراد متحررة من أثر الفقرات (Item-Free Person).

النظرية الحديثة في القياس:

تعد النظرية الحديثة في القياس ثورة في القياس النفسي والتربوي، وذلك كما يعتبرها كثير من علماء القياس (علام، 1986). تفترض هذه النظرية أنه يمكن التنبؤ بأداء الأفراد في ضوء خاصية أو خصائص مميزة لهذا الأداء تسمى السمة (trait) بحيث تحاول هذه النظرية تقدير علامات الأفراد على هذه السمة التي تصعب ملاحظتها مباشرة، لـذا تقـدر مـن أداء الأفراد على مجموعة فقرات الاختبار.

نتجت عن هذه النظرية مجموعة من النماذج تعرف باسم نمـاذج الـسمات الكامنـة (latent trait model) تهدف إلى تحديد العلاقة بين أداء الفرد في الاختبار وبين الـسمة أو القدرات التي تكمن وراء هذا الأداء وتفسره.

وقد وضع ماكدونالد (McDonald, cited in Hambleton, 1989) إطاراً عامـاً ليس فقط لتنظيم هذه النماذج، وإنما لتوليد العديد من النماذج الجديدة، حيث يتـضمن إطـاره الاهتمام بـــ:

1. نماذج أُحادية البعد ومتعددة الأبعاد (Unidimensional & Multidimensional Models).
2. نماذج ثنائية ومتعددة الاستجابة (Dichotomous & Polytomous Response model).
3. نماذج خطية وغير خطية (Linear & Nonlinear Models).

لكن إلى اليوم، الاهتمام الأكبر موجه نحو نماذج الاستجابة الثنائية ذات البعد الواحد، غير الخطي.

وترتكز نماذج النظرية الحديثة للقياس أو نظرية استجابة الفقرة (Item Response Theory, (IRT))، على مجموعة من الافتراضات التي يجب أن تتوافر في البيانات المستمدة من الاختبار، حيث يعتمد الاختيار الصائب للنموذج على تحقيق هذه الافتراضات في البيانات.

الافتراضات التي تتميّز بها نماذج السمات الكامنة:

أي نموذج رياضي يتضمن مجموعة افتراضات حول البيانات التـي سيطبق فيهـا النموذج. هناك أربعة افتراضات لنماذج استجابة الفقرة هي:

أولاً: أُحادية البعد (Unidimensionality): أي وجود قدرة واحدة تفسر أداء الفرد فـي الاختبار.

ثانياً: الاستقلال الموضعي (Local Independence) الاستقلال الموضعي للفقرات على أي نقطة معطاة على متصل السمة أي تكون علامات الفقرة مستقلة الواحدة عن الأخرى.

ثالثاً: وجود اقتران رياضي يربط بين احتمال نجاح الفرد في الإجابة عن فقرة اختبار، وبين القدرة التي تقيسها مجموعة الفقرات التي يشتمل عليها هذا الاختبار، ويعـرف التمثيـل البياني لهذا الاقتران باسم منحنى خاصية الفقرة (Item Characteristic Curve,(ICC)).

رابعاً: عامل السرعة (Speededness) حيث تفترض نماذج (IRT) أن عامل السرعة لا يلعب دوراً في الإجابة عن فقرات الاختبار، بمعنى أن إخفاق الأفراد في الاختبار يرجع إلى انخفاض قدرتهم وليس إلـى تـأثير عامـل السـرعة علـى الإجابـة (Hambleton & Swaminathan, 1985).

نماذج السمات الكامنة:

أ- النموذج التجمعي الطبيعي (Normal-ogive model).

ب- النموذج اللوجستي أُحادي المعلَمة (One-parameter logistic model).

ج- النموذج اللوجستي ثنائي المعلَمة (Two-parameter logistic model).

د- النموذج اللوجستي ثلاثي المعلَمة (Three-parameter logistic model).

حيث تختلف هذه النماذج في الصورة الرياضية التي تمثل المنحنى المميز للفقرة.

أولاً: النموذج اللوجستي أُحادي المعلَمة: One Parameter Logistic Model

يفترض هذا النموذج أن جميع الفقرات تميّز بنفس القدر بين الأفراد، لكنها تتباين فقـط في صعوبتها.

ومعادلته الرياضية هي:

$$P_i(\theta) = \frac{1}{1+e^{-Da(\theta-bi)}} \qquad (1)$$

حيث $P_i(\theta)$: احتمال أن يجيب المفحوص ذو القدرة (θ) على الفقـرة (i) إجابـة صحيحة.

D: عامل التدريج (Scaling factor) وهو ثابت.

a: معامل تمييز الفقرة (هنا ثابت لجميع الفقرات)

bi: معامل الصعوبة للفقرة (i).

ثانياً: النموذج اللوجستي ثنائي المعلَمة:

توصل بيرنبوم (Birnbaum, cited in Hambleton, 1989) إلى هذا النموذج حيث يتخذ المنحنى المميز للفقرة في هذا النموذج المعادلة الآتية:

$$P_i(\theta) = \frac{1}{1+e^{-Dai(\theta-bi)}} \qquad (2)$$

يفترض هذا النموذج والنموذج السابق (أحادي المعلَمة) عدم تـأثر الإجابـات بعامـل التخمين (عامل الصدفة).

ويتميّز منحنى هذه العلاقة بأنه:

1. أملس (ممهمد) Smooth.
2. يتزايد بزيادة θ.
3. له خط تقارب أفقي (Horizontal Asymptotes) عند القيم القصوى لــ θ.

ثالثاً: النموذج اللوجستي ثلاثي المعلَمة:

يعتبر هذا النموذج امتداداً للنموذج السابق، مضيفاً إليه معلَماً جديداً يسمى الخط التقاربي لأسفل المنحنى المميز للفقرة (Lower Asymptote) الذي يمثل احتمال توصل الأفراد ذوي القدرة المنخفضة إلى الإجابة الصحيحة عن الفقرة.

والمعادلة الرياضية لهذا المنحنى هي:

$$P_i(\theta) = c_i + (1 - c_i)\frac{e^{Dai(\theta - bi)}}{1 + e^{Dai(\theta - bi)}}\ ; i = 1, 2, \ldots, n \qquad (3)$$

(Hambleton, 1989)

بناء الاختبارات وفق النظرية الحديثة في القياس:

إن عملية تطوير وبناء الاختبارات سواء بالطريقة التقليدية أم بالطريقة الحديثة تمـــر بتسع خطوات رئيسة هي:

1. تحديد الغرض من الاختبار.
2. تحضير تجمع كبير (Pool) من الفقرات.
3. فحص ميداني للفقرات (جمع بيانات وتحليلها).
4. اختيار فقرات الاختبار.
5. إخراج الاختبار بصورته النهائية.
6. تجميع المعايير (Compilation of Norms).
7. تحديد علامة القطع (للاختبارات محكية المرجع).
8. دراسة الثبات.
9. دراسة الصدق (Hambleton, 1989).

إن الفروق الرئيسة في بناء الاختبار باستخدام الطريقتين: الحديثة والتقليدية تحدث عند الخطوات 3، 4، 6، 8.

وفيما يتعلق بالفحص الميداني (Field testing) للفقرات، فإن الحسنة الكبرى لنظرية استجابة الفقرة هي أنها تقود إلى فقرات تتسم معالمها باللاتغير عند تغيير العينة، لكن الصعوبة تكمن في كبر حجم العينة المطلوبة للحصول على تقديرات مستقرة لمعالم الفقرة بالإضافة إلى التعقيدات الرياضية للأساليب المستخدمة للحصول على هذه التقديرات.

وإن اكتشاف الفقرات غير الملائمة لا يكون مباشراً كما هو الحال في أساليب النظرية التقليدية، حيث تقوّم الفقرات بالنسبة إلى مدى ملاءمتها للنموذج المستخدم باستخدام اختبار إحصائي أو تحليل البواقي.

إن الطريقتين (التقليدية والحديثة) تتفقان في تحليل مموهات الفقرة لاكتشاف فعالية الفقرة.

أما عند اختيار الفقرات (Item Selection) الخطوة الرابعة، فتعرض النظرية الحديثة للقياس لباني الاختبار طريقة أخرى لاختيار الفقرة، ولكن عملية الاختيار للفقرة، هي كالطريقة التقليدية معتمدة على غرض الاختبار، إلا أن اختيار الفقرات يعتمد على كمية المعلومات التي تساهم بها الفقرة إلى المقدار الكلي من المعلومات الذي يزود به الاختبار حيث إن استخدام دالة معلومات الفقرة تتيح لباني الاختبار إنتاج اختبار يتسم بالدقة، وبتوفير فرصة الحصول على تقديرات دقيقة لمعلم القدرة وخصوصاً بالنسبة للاختبارات محكية المرجع.

وهناك طرق عديدة لتقدير معلم القدرة ومعالم الفقرات منها طريقة:

- الأرجحية العظمى للتقدير ((Maximum Likelihood Estimation (MLE)).
- الأرجحية العظمى المشروطة للتقدير (Conditional Maximum Likelihood Estimation).

- الأرجحية العظمى الهامشية ((Marginal Maximum Likelihood Estimation (MLE)).
- بييز للتقدير (Bayesian Estimation).

وتقدم برمجية Bilog - Mg تقديرات للقدرة وفق الطريقتين الأرجحية العظمى الهامشية وبييز، اعتمدت الباحثة طريقة الأرجحية العظمى الهامشية (MML) والتي تعمل على إيجاد θ التي تجعل قيمة (θ) Log أكبر ما يمكن من خلال المعادلة الآتية:

$$logl_i(\theta) = \sum_{j=1}^{n} \{X_{ij} \log_e P_j(\theta) + (1 - X_{ij}) \log_e [1 - P_j(\theta)]\} \quad (4)$$

حيث يشير الرمز $P_j(\theta)$ إلى دالة الاستجابة للفقرة j. ولإيجاد أكبر قيمة لـ θ نجد المشتقة الأولى للمعادلة السابقة (4) ونساويها بالصفر، ثم نحل المعادلة:

$$\frac{\partial \log Li(\theta)}{\partial \theta} = \sum_{j=1}^{n} \frac{Xij - Pj(\theta)}{P_j(\theta)[1 - P_j(\theta)]} \cdot \frac{\partial Pj(\theta)}{\partial \theta} = 0$$

يتم حساب تقديرات الأرجحية العظمى (*Ml*) باستخدام طريقة تدريج فيشر Fisher- Scoring Method والتي تعتمد على ما يسمى بمعلومات فيشر Fisher-Informationالتي تعطى بالمعادلة الآتية.

$$I(\theta) = \sum_{j=1}^{n} a_j^2 Pj(\theta)[1 - Pj(\theta)] \quad (5)$$

في حالة النموذج ثنائي المعلَمة وهناك صيغ مشابهة في حالة النماذج الأخرى.

يتم إيجاد الخطأ المعياري لتقديرات الأرجحية العظمى من خلال المعادلة:

$$S.E(\hat{\theta}) = \sqrt{1 / I(\hat{\theta})} \quad (6)$$

كما تستخدم هذه الطريقة لتقدير معالم الفقرات للاختبار وفق النماذج الثلاثة، حيث تتميّز هذه الطريقة بأنها تعطي تقديرات دقيقة عن طريقة إعادة متعاقبة لعمليات التقدير، وتتميّز هذه الطريقة بأنها تطبق لكل نماذج IRT، بالإضافة إلى الكفاءة والفاعلية التي تمتاز

بها لكل من الاختبارات القصيرة والطويلة. وتعطي تقديرات للعلامة الكلية وبالتالي لا يوجد فقدان للمعلومات عائد بسبب حذف استجابة بعض أفراد الدراسة، بالإضافة إلى أن تقديرات هذه الطريقة متسقة وتقترب من القيم الحقيقية بزيادة حجم العينة.

وأما بالنسبة إلى دراسة الثبات (Test Reliability) يتم البحث في دالة المعلومات للاختبار كمقياس لدقة التقدير (Accuracy of Estimation) حيث إن شكل الدالة يعتمد فقط على الفقرات المتضمنة في الاختبار ويزود بتقدير لخطأ القياس عند كل مستوى للقدرة.

وتزود بعض البرامج الحاسوبية بطرق استخراج الثبات، فمن خلال برمجية Bilog-Mg التي اعتمدتها الباحثة لتحليل بياناتها، تم تعريف الثبات بالطريقتين الآتيتين:

1. الثبات النظري (Theoretical Reliability):

تعتمد هذه الطريقة في تقدير قيمة الثبات على تقديرات نظرية استجابة الفقرة لمعالم الفقرات المستخرجة من خلال طريقة الأرجحية العظمى (Maximum Likelihood Method (ML)، حيث تفترض هذه الطريقة أن العلامات الحقيقية للقدرة تتوزع توزيعاً طبيعياً بمتوسط حسابي مقداره صفر وانحراف معياري يساوي واحد، تحسب دالة معلومات الفقرة من خلال حساب الخطأ المعياري للقياس حيث دالة معلومات الفقرة تعطى بالعلاقة.

$$I(\theta) = \frac{1}{S.E^{2}(\theta)} \qquad (7)$$

وأما الخطأ المعياري $S.E$ فيعطى لكل نموذج من خلال المعادلات الآتية:

- **النموذج أحادي المعلَمة:**

$$S.E_{(1)}(\hat{\theta}) = \left\{1/D^2a^2\sum_{j=1}^{n}P_{(1)j}(\hat{\theta})\left[1-P_{(1)j}(\hat{\theta})\right]\right\}^{\frac{1}{2}} \qquad (8)$$

- **النموذج ثنائي المعلَمة:**

$$S.E_{(2)}(\hat{\theta}) = \left\{1/D^2\sum_{j=1}^{n}a^2_jP_{(2)j}(\hat{\theta})\left[1-P_{(2)j}(\hat{\theta})\right]\right\}^{\frac{1}{2}} \qquad (9)$$

- النموذج ثلاثي المعلَمة:

$$S.E._{(3)}(\hat{\theta}) = \left\{ 1/D^2 \sum_{j=1}^{n} a^2 \frac{1-P_{(3)j}(\hat{\theta})}{P_{(3)j}(\hat{\theta})} \left\{ \frac{P_{(3)j}(\hat{\theta}) - gj}{1-gj} \right\}^2 \right\}^{\frac{1}{2}} \quad (10)$$

يتم حساب الوسط الحسابي لتباين الخطأ من خلال إيجاد مقلوب الوسط الحسابي لدالة المعلومات، وبالاعتماد على افتراض أن تباين العلامات الحقيقية للتوزيع يساوي واحداً، فإن الثبات النظري يساوي واحداً مقسوماً على تباين الخطأ مضافاً إليه العدد واحد.

2. **الثبات الأمبيريقي (التجريبي) (Empirical Reliability):**

يعتمد استخراج الثبات الأمبيريقي على كيفية تقدير علامات القدرة للمفحوصين في العينة، حيث يقدم البرنامج طريقتين الأولى الطريقة المعتمدة على علامات الأرجحية العظمى (ML) والطريقة الثانية المعتمدة على علامات بييز Bayes.

واختارت الباحثة الطريقة الأولى الأرجحية العظمى حيث يستخرج الثبات التجريبي للعلامات من خلال إيجاد ناتج قسمة تباين العلامة الحقيقية على تباين العلامة.حيث تباين العلامة هو تباين العلامات المستخرج من خلال الأرجحية العظمى، أما التباين الحقيقي فيتم تقديره عن طريق طرح تباين الخطأ من تباين العلامة.

وفيما يتعلق بالاتساق الداخلي يمكن استخراجه عن طريق:

معادلة ثبات الاتساق الداخلي (كودر ريتشاردسون 20) (KR-20):

يتم استخراج (KR-20) من خلال برمجية SPSS وفق المعادلة الآتية:

$$P_{xx'} = \left(\frac{n}{n-1}\right)\left(1 - \frac{\sum_{i=1}^{n} Pi(1-Pi)}{S^2}\right) \quad (11)$$

حيث Pi = معامل الصعوبة.

n = عدد سالفقرات.

s^2= التباين للعلامات على الاختبار.

ولقد حدد لورد (Lord, 1980) خطوات لتطوير وبناء الاختبار تساعد في استلال اختبار من تجمع فقرات (item pool) تمت معايرتها سابقاً، يتسم هذا الاختبار بالدقة ويحقق الخصائص التي ينشدها باني الاختبار، هذه الخطوات هي:

1. حدد شكل دالة معلومات الاختبار الذي تريد (Target information curve).
2. اختر الفقرات التي تعمل على الاقتراب من الشكل المطلوب لدالة المعلومات.
3. أضف الفقرات فقرة تلو الأخرى، فاحصاً في كل مرة شكل دالة المعلومات.
4. استمر في اختيار الفقرات وإضافتها حتى تحصل على شكل الدالة المطلوب للمعلومات.

بهذه الخطوات يتم الحصول على اختبار يميز جيداً مما يعمل على زيادة دقة تقديرات المفحوصين.

الدقة النسبية والفاعلية النسبية (Relative Precision & Relative Efficiency):

غالباً ما تكون هناك حاجة لمقارنة كمية المعلومات الناتجة من قبل نماذج مختلفة للاختبار، إلا أن التفسيرات المطلقة لا يمكن الحصول عليها مباشرة من دالة معلومات الفقرة أو الاختبار، أو حتى العلامة الخام. إلا أنه يمكن الاستفادة من دالة معلومات الاختبار بحساب نسبة دالة المعلومات موضع الاهتمام حيث تسمى هذه النسبة بالدقة النسبية(Relative Precision) ويرمز لها بالرمز RP(θ) (Hambleton & Swaminathan,1985).

$$R\,P\,(\theta) = I_1\,(\theta)\,/\,I_2\,(\theta) \qquad (12)$$

حيث يشير الرمزان I_2, I_1 إلى دالتي معلومات نموذجي الاختبارين 1,2.

وفي الحالات الخاصة التي يتم فيها مقارنة دالتي معلومات لاختبارين من نفس النموذج تسمى النسبة في هذه الحالة بالفاعلية النسبية (Relative Efficiency) ويرمز لها بالرمز $RE(\theta)$ حيث:

$$RE(\theta)= I_1 (\theta) / I_2 (\theta) \qquad (13)$$

وتستخدم الفاعلية النسبية لتقييم الأثر الناتج من تغيير بعض الفقرات في الاختبار الأصلي حيث يتم حساب الفاعلية النسبية للاختبارين: الأصلي والمنقح (Lord, 1980).

ومع أن دوال المعلومات تتأثر بشكل كبير بالتحويلات على مقياس القدرة، فإن دالة الفاعلية النسبية لا تتأثر بأي تحويل على مقياس القدرة، وإذا كان لدى الباحث سبب هام لاستخدامه مقياس قدرة خاص، فإنه من المعقول أن يقوّم الاختبار بدلالة دالة المعلومات، وفي كثير من الحالات فإن دالة الفاعلية النسبية تفضل أكثر (Allen & Yen, 1979).

مشكلة الدراسة:

تنشأ الحاجة إلى اختبارات التفكير الناقد من أهمية التفكير الناقد نفسه، الذي زاد الاهتمام به في السنوات الأخيرة من القرن الماضي، ويتضح من العرض السابق (الوارد في بند حركة قياس التفكير الناقد من الرسالة نفسها) أن هناك حاجة ماسة لاختبارات التفكير الناقد، وقد زادت أهمية التفكير الناقد في السنوات الأخيرة من القرن الماضي، وتعد مهارات التفكير الناقد عنصراً هاماً في التعليم المدرسي والجامعي.

من هنا سعت هذه الدراسة بشكل رئيسي إلى بناء اختبار للتفكير الناقد وفق نماذج النظرية الحديثة في القياس "نظرية استجابة الفقرة"، يضم كافة المهارات التي ضمنتها المقاييس السابقة للتفكير الناقد مثل مهارة التحليل والتفسير والتقويم، وإضافة مهارة التبرير الإحصائي

والاحتمالي إلى المقياس الحالي والتي لم ترد في أي مقياس سابق، من خلال بناء فقرات تتضمن مواقف حياتية تضع المفحوص أمام مشكلة يتطلب حلها استخدام إحدى مهارات التفكير الناقد.

واختارت الباحثة الفئة العمرية بدءاً من العمر 15 سنة لأن هذه الفئة وحسب تصنيف بياجيه يكون الفرد فيها في مرحلة العمليات الرسمية (Formal Operations) التي تبدأ من سن 11 سنة حتى الرشد، حيث تتميّز هذه المرحلة بقدرة الفرد على تبرير الأفكار المجردة، الافتراضية، العكسية، ويتمكن من صياغة واختبار الفرضيات المتعددة والمهتمة بربط السبب مع النتيجة، ويستطيع أيضاً في هذه المرحلة من فصل وضبط المتغيرات.

وتحديداً فإن الدراسة سعت إلى تحقيق الأهداف الآتية:

1. بناء اختبار للتفكير الناقد وفق نماذج نظرية استجابة الفقرة للقياس للفئة العمرية 15 سنة فما فوق.
2. دراسة دالة المعلومات للاختبار وفق النماذج الثلاثة لنظرية استجابة الفقرة، وتقدير الدقة النسبية لها.
3. دراسة الفاعلية النسبية لصورة مصغرة من الاختبار وفق النماذج الثلاثة المسحوبة منها.
4. تحديد المعايير (Norms) والمئينية بالإضافة إلى تدريج القدرة.

ولتحقيق تلك الأهداف فإن الدراسة سعت للإجابة عن الأسئلة الآتية:

1. ما درجة مطابقة الاستجابات عن فقرات الاختبار لنماذج نظرية استجابة الفقرة؟
2. ما قيم تقديرات معالم الفقرة والقدرة على الاختبار وفق النماذج الثلاثة لنظرية استجابة الفقرة؟
3. ما الدقة النسبية للاختبار وفق النماذج الثلاثة لنظرية استجابة الفقرة؟

4. ما الفاعلية النسبية للاختبارات الجزئية المستلة من الاختبارات الكلية؟

5. ما دلالات صدق الاختبار؟

6. ما دلالات ثبات الاختبار؟

7. ما قيم المعايير المئينية؟

أهمية الدراسة:

يرى باير (Beyer, 1985) أن تعلم التفكير الناقد يعزز فرص المجتمعات والأفراد على البقاء في ظل التغيرات المتسارعة في هذا العالم، ومن هذا المنطلق نادى الباحثون والتربويون لسنوات عديدة بضرورة تعليم مهارات التفكير الناقد للطلبة، واختلف الباحثون ما بين تضمين تلك المهارات وتعليمها من خلال المحتوى الدراسي للمواد المختلفة (Gadzella, Hartsoe & Harper, 1989) أو تدريسها بشكل مباشر ومنظم من خلال مساق منفصل يقوم بتعليم مهارات التفكير الناقد (Brookfield, 1997; Halpern, 2001; Paul & Elder, 2004).

ولما كانت مهارات التفكير الناقد في مثل هذه الأهمية فإن من الضرورة العمل على تطوير الأدوات اللازمة والمناسبة لقياس هذه المهارة، والعمل على تطويرها وتنميتها، من هنا تكمن أهمية هذه الدراسة في توفير أداة لقياس مهارات التفكير الناقد مبنية وفق أسس القياس الموضوعي الذي يتمتع بفقرات ذات خصائص سيكومترية متحررة من الأفراد، وبتقدير قدرات الأفراد بصورة موضوعية متحررة عن الفقرات التي يأخذونها.

كذلك من الممكن سحب عينات عشوائية من الفقرات من المجال الكلي (Universe Domain) لهذه الفقرات واختبار كل فرد بإحدى هذه العينات، وتقدير قدرة كل منهم على تدريج مشترك.

ونظراً لأن معالم الفقرات لا تتغير (Invariant) بتغير مجموعة الأفراد التي تختبر بها، فإن بناء الاختبار وفق النظرية الحديثة للقياس يمكننا من التمييز بين الأفراد عند مختلف مستويات القدرة.

وتبرز أهمية هذه الدراسة أيضاً من أهمية موضوع التفكير الناقد نفسه في جميع مجالات الحياة في المجتمعات الحديثة، إذ يعده الكثير من الباحثين بأنه حق لكل متعلم ومتطلب رئيسي لجميع أفراد المجتمع (Paul, 1998).

وتكمن أهمية هذه الدراسة أيضاً من المدى الواسع من المهارات الفرعية التي يغطيها هذا الاختبار ويعمل على قياسها، حيث يوفر قياساً أكثر دقة وصدقاً للتفكير الناقد كما يفيد في معرفة مستوى التطور في مهارة التفكير الناقد عند الفرد، وتشخيص جوانب التفوق أو الضعف في تطور هذه المهارة، وبيان كمية ما يمتلك منها قبل البدء بالتدرب عليها، وبالتالي يسهل اتخاذ القرارات التربوية، والإدارية التي تتعلق بتخطيط البرامج لتنمية تلك المهارة.

تعريف المصطلحات:

- **التفكير الناقد**: أسلوب في التفكير المنظم والمركز والمبرر بغرض حل المسألة وصياغة الاستدلالات وحساب الاحتمالات وعمل القرارات، ويقاس من خلال الاختبار الذي أعد لأغراض هذه الدراسة والذي يتضمن المهارات الآتية.

1. **التحليل (Analysis)**: تحديد العلاقات ذات الدلالات المقصودة والفعلية بين العبارات، والأسئلة والمفاهيم والصفات، والصيغ الأخرى للتعبير عن اعتقاد أو حكم أو تجربة أو معلومات أو آراء (Facione & Facione, 1998).

1.أ. فحص الآراء (Facione & Facione, 1998).

1.ب. اكتشاف وتحليل الحجج (Facione & Facione, 1998).

2. **التفسير (Interpretation):** الفهم والتعبير عن المعنى أو الدلالة لمختلف التجارب والمواقف والمعطيات والقوانين والإجراءات والمعايير (Facione & Facione, 1998).

2.أ. التصنيف (Facione & Facione, 1998).

2.ب. استخلاص المغزى (Facione & Facione, 1998).

2.ج. توضيح المعاني (Facione & Facione, 1998).

3. **التقويم (Evaluation):** تقييم مصداقية وصحة العبارات التي تصف أو تفسر فهم أو إدراك الشخص لتجاربه ومعتقداته وآرائه، بالإضافة إلى تقييم القوة المنطقية للعلاقات ذات الدلالة الحقيقية بين العبارات والصفات والأسئلة أو أي شكل آخر من أشكال التعبير (Facione & Facione, 1998).

3.أ. تقويم الإدعاء (Facione & Facione, 1998).

3.ب. تقويم الحجج (Facione & Facione, 1998).

3.ج. تقويم الحلول (Scriven & Paul, 1992).

4. **الشرح (Explanation):** قدرة الفرد على إعلان نتائج تفكيره وتبرير ذلك التفكير بالأدلة والمفاهيم والمنهجية والمعايير المنطقية والقرائن، وعرض تبريره أو تفكيره على شكل حجج مقنعة (Facione & Facione, 1998).

4.أ. وضع النتائج (Facione & Facione, 1998).

4.ب. تبرير الإجراءات (Facione & Facione, 1998).

4.ج. تقديم الحجج (Facione & Facione, 1998).

5. **الاستدلال (Inference)**: تحديد العناصر اللازمة لاستخلاص نتائج معقولة، وتكوين تخمينات وفرضيات، والاهتمام بالمعلومات ذات الصلة، واستخلاص ما يترتب من البيانات والعبارات والأسئلة أو أي شكل آخر للتعبير، بالإضافة إلى فحص واختبار القدرة على التمييز في درجة صحة أو خطأ الاستدلال المشتق من بيانات معطاة (Facione & Facione, 1998; Watson & Glaser, 1964).

5.أ. تقصي الأدلة (Facione & Facione, 1998).

5.ب. تخمين البدائل (Facione & Facione, 1998).

5.ج. التوصل إلى الاستنتاج (Facione & Facione, 1998; Watson & Glaser, 1964).

5.د. فحص أو اختبار القدرة على التمييز في درجة صحة أو خطأ الاستدلال المشتق من بيانات معطاة (Beyer, 1985; Ennis, 1985; Watson & Glaser, 1964).

6. **التبرير الاحتمالي والإحصائي** (Halpern, 2001; Sawin, 2004):

6.أ. استخلاص الفكرة أو المغزى من البيانات أو الرسم.

6.ب. تفسير البيانات أو الرسم.

7. **التعرّف إلى الافتراضات المنصوص وغير المنصوص عليها**: اختبار أو فحص القدرة على التعرف إلى الافتراضات المنصوص وغير المنصوص عليها في جملة أو نص معين، والتي تم التسليم بصحتها (Beyer, 1985; Ennis, 1985).

8. **النظر إلى الأشياء بعمق وتوسع (التركيز)** (Ennis, 1985):

8.أ. النظر إلى الأشياء بمقدار أوسع.

8.ب. النظر إلى الأشياء بعمق أكبر.

الفصل الثاني

الدراسات السابقة

كثر الحديث في الآونة الأخيرة حول التفكير بشكل عام، والتفكير الناقد بشكل خاص، مما دفع بالساسة قبل التربويين أن يولوه أهمية كبرى وتشعبت الميادين التي بحثت فيه، فمن ميدان التمريض، إلى ميدان الجيش، إلى ميدان الإدارة، فالجامعات فالمدارس. وبالتالي كثرت الدراسات التي تناولته سواء لوحده أم مع متغيرات أخرى (Alter & Egan, 1997; Browne & Keeley, 1988; McCarthy, 2004; Sormunen & Chalupa, 1994).

تمت مراجعة واستعراض الكثير من الدراسات التي اهتمت بموضوع التفكير الناقد. ويلاحظ أن الكثير من تلك الدراسات قد تناولت استراتيجيات تنمية مهارات التفكير الناقد وعلاقة التفكير الناقد ببعض المتغيرات مثل (الجنس، والمستوى التعليمي، والتحصيل) (بشارة، 2003؛ جبريل وجوابرة، 2003؛ حمادنة، 1995؛ الخطيب، 1993؛ خليفة، 1990؛ الخوالدة، 2002؛ الدردور، 2001؛ الربضي، 2004؛ الشبول، 2004؛ الصويتي، 2001؛ العبد اللات، 2003؛ Astleitner, 2002؛ Burbach, Matkin & Fritz, 2004؛ Dean, 2004; Gadzella, Hartsoe and Harper, 1989; Schuck, 2003).

ولقد تركز اهتمام الباحثة في هذا الفصل بمراجعة واستعراض الدراسات التي اهتمت بقياس التفكير الناقد، وبالخصائص السيكومترية لمقاييسه العالمية:

أجرى كل من مودجسكي وميشيل (Modjeski & Michael, 1983) دراسة هدفت إلى تقييم الصدق والثبات والخطأ المعياري في القياس لاختبارين من اختبارات التفكير الناقد هما اختبار كورنل- المستوى X (Cornell Critical Thinking Test- Level X) واختبار واتسون -جليسر النموذج YM (Watson & Glaser-Form YM). وتم اعتماد عشرة معايير لتقييم الصدق وخمسة معايير لتقييم الثبات وأخطاء القياس استمدت من المعايير العامة للاختبارات التربوية والنفسية (Publication Standards for Educational & Psychological Tests). تكونت لجنة التقييم (التحكيم) من 12 عالمَ نفس من حملة درجة الدكتوراه وممن قد نُشرت لكل منهم مقالة على الأقل في مجال التفكير الناقد في دورية محكمة (Refereed Journal). يشغل أحد عشر محكماً منهم منصباً أكاديمياً في جامعة كبرى. طلب من كل محكم إصدار حكم لكل معيار من المعايير التي تم اعتمادها، بحيث يقوم باختيار التدريج المناسب لمدى تحقق المعيار من بين بشكل تام، قليل، قليل جداً، لم يتحقق، والتي تعطى القيم التدريجية 4 ،3، 2، 1 على الترتيب. أظهرت نتائج التحليل أن التقييمات في الغالب ما تكون تحقق المعيار بشكل قليل أو قليل جداً. بالإضافة إلى عدم تحقق معيار الاستقرار في العلامات (Stability of Test Score) لكلا الاختبارين، وكان هناك تباين في تقديرات المحكمين لمدى تحقق معايير الصدق والثبات في كلا الاختبارين، وتفوّق اختبار واتسون – جليسر على اختبار كورنل في عدد المعايير التي تحققت في كل من الصدق والثبات، لكن تبين ملائمة العينة المستخدمة لاستخراج معامل الثبات وأخطاء القياس.

وفي دراسة أجراها فرسبي (Frisby, 1991) هدفت إلى دراسة العلاقة بين مستوى الصف، والمتوسطات الحسابية لأداء الطلبة على اختبار كورنل للتفكير الناقد (النموذج x) من خلال تحليل مجموعة من الدراسات السابقة. تكونت عينة الدراسة من 29 دراسة تم تنفيذها ما بين 1960 إلى 1985، كانت هذه العينات لطلبة من ذوي المستويات المتوسطة أو الموهوبة من الصفوف 4-12 بالإضافة إلى سنة أولى جامعة ومجموعة من معلمي المدارس.

تناولت تلك الدراسات مجموعات متباينة من الطلبة في الصف والعمر. أشارت النتائج إلى: نقصان مقدار الفروق بين متوسطات الأداء في العينات ذات الأعمار الأكبر، وأن الاختبارات تميل إلى إعطاء تقديرات أعلى للمفحوصين بفعل عامل التخمين الناتج من فقرات الاختبار من نوع الاختيار من متعدد، وهناك زيادة كبرى في حجم الأثر متصاحبة مع التغير من الفئة الصفية السابع والتاسع إلى المرحلة الثانوية العاشر والثاني عشر.

وفي دراسة أجراها كل من بول ونوسج (Paul & Nosich, 1992) بهدف البحث عن أداة التقييم المناسبة لمهارات التفكير الناقد، حيث قاما بدراسة اختبارات التفكير الناقد الموجودة عالمياً كاختبار كورنل وروس واختبار مهارات استقصاء ونيوجرسي حيث أفرزت

المراجعة على أنه لم يلعب تعريف التفكير الناقد دوراً أساسياً في تصميم أدوات التقييم لغاية الآن، لأن المفهوم قد تم تطويره بشكل شامل فقط خلال السنوات العشر الأخيرة، ولم يتوفر الوقت الكافي بعد لتطوير أدوات تقييم جديدة، ويفضل أن تحتوي الفقرات أمثلة حياتية حقيقية قدر الإمكان، ومسائل وقضايا تشغل تفكير الناس، كما يفضل أن يحتوي الاختبار على مجال واسع من القدرات والمهارات الأساسية المتكاملة، وهناك قصور في كل اختبارات التفكير الموجودة، حيث تفتقر إلى بعض المعايير التي ينبغي أن تتوفر في الاختبارات التي تقيس مهارات التفكير العليا، ولم يمثل أي من الاختبارات الموجودة حالياً أنموذجاً شاملاً لعناصر التفكير الناقد.

وفي دراسة أجراها جاكوبس (Jacobs, 1995) هدفت إلى دراسة خصائص اختبار كاليفورنيا النموذجين A ,B. تكونت عينة الدراسة من 1646 طالب وطالبة من طلبة السنة الأولى في جامعة (Large Eastern Private University) عام 1993. كانت نسبة الذكور 52%. طبق النموذج A، على 684 طالب وطالبة في حين تم تطبيق النموذج B من نفس الاختبار على 962 طالب وطالبة. أظهرت نتائج التحليل الإحصائي أن هناك اختلافات في صعوبة الفقرة بين النموذجين، وضعف الاتساق الداخلي بين الفقرات، وعدم تحقق خاصية أُحادية البعد لفقرات الاختبار، وأن العلامة الكلية على الاختبار تبدو أنها مناسبة وكافية فقط لأغراض بحثية، وليس لأغراض صنع القرارات، وبغض النظر عما يقيسه النموذجان فإن مقارنة علامتيهما معاً موضع تساؤل كبير، وإن تحليل الاختبارات الفرعية للنموذجين A, B تشير إلى أن بعض الاختبارات الفرعية أكثر تجانساً، وبذلك أكثر قابلية للتفسير من العلامة الكلية (مثل الاختبار الفرعي لمهارة الاستنتاج Deduction)، بالإضافة إلى أن فشل التحليل في إظهار العناصر الأساسية لتصنيف الفقرات كما اقترحه الاختبار والاختلاف بين النموذجين

يجعل استعماله موضع تساؤل كبير. وأوصت الدراسة بإجراء المزيد من الدراسات التي تساعد في تحديد مدى صلاحية علامات الاختبار في قياس السمة، ولأي مدى تكون هذه الأداة حساسة للتغيرات التي تحدث نتيجة برامج مصممة لتنمية مهارات التفكير الناقد.

وأجرت السيد (1995) دراسة هدفت إلى: بناء مقياس للتفكير الناقد للأطفال في المرحلة العمرية من (9–12) سنة، وتحديد علاقته بكل من متغيرات الجنس، والحياة الثقافية في المجتمع، ومتغير المناخ المدرسي، ومتغير المرحلة العمرية. وتكونت عينة الدراسة من 247 طالب وطالبة موزعة على ثلاث مجموعات: المجموعة الأولى تتكون من طلبة الصفوف: الرابع، والخامس، والسابع من مدرستين تابعتين للمدارس الخاصة في جمهورية مصر العربية بلغ مجموعهم 80 طالباً من بينهم 55 طالبة، والمجموعة الثانية تتكون من طلبة الصفوف: الخامس، والسابع، من أربع مدارس حكومية، بلغ عددهم 48 طالباً وطالبة نصفهم من الطالبات في جمهورية مصر العربية، والمجموعة الثالثة تتكون من طلبة الصفوف: الخامس، والسادس، والسابع، من المدارس في دولة قطر، بلغ مجموعهم 119 طالباً وطالبة منهم 72 طالبة. حيث قامت الباحثة ببناء أداة لقياس التفكير الناقد لدى أطفال المرحلة العمرية من سن التاسعة وحتى سن الثانية عشرة. تكوّن المقياس من 85 فقرة تقيس المهارات الآتية الدقة في فحص الوقائع (22 فقرة)، والاستدلال(26 فقرة)، والقدرة على الاستنتاج (25 فقرة)، وتقويم الحجج والمناقشات (12 فقرة)، وأسفرت نتائج الدراسة عن وجود فروق تعزى للثقافة السائدة ونوع المجتمع والمناخ المدرسي، عللت الباحثة هذه النتيجة إلى الاختلاف في الأنشطة ودور الأسرة، وشعور الطفل بأهميته، فضلاً عن المستوى الاقتصادي والاجتماعي. كما أسفرت الدراسة عن وجود فروق بمهارات التفكير الناقد بين الأطفال الصغار والكبار ولصالح الكبار بينما لا يوجد فروق تعزى لمتغير الجنس.

وأجرى بايك (Pike, 1997) دراسة هدفت إلى دراسة الخصائص السيكومترية لاختبار كاليفورنيا، عرّف من خلالها الاختبار والمراحل التطويرية التي مر بها والمهارات التي يقيسها، كما قارن بين الخصائص السيكومترية للنموذجين A ,B موضحاً أن النموذج B هو نموذج مطوّر من النموذج A، تم فيه إعادة صياغة 28 فقرة من أصل 34 فقرة موجودة في النموذج A، شملت التغيرات: الاختلاف في الأسماء والمفاهيم والسياق، ونوع المشكلات المطروحة في الفقرة، لكن لم يطرأ تغيير على المهارات التي قام بقياسها النموذج A، وأشار الباحث إلى أن تقديرات الثبات التي عرضها فاسيون (Facione) تراوحت بين (0.68 – 0.69). في حين كانت تقديرات الثبات في دراسة جاكوبس (Jacobs, 1995) للنموذجين A، B 0.56 و 0.59 على الترتيب. وباستخدام معامل الثبات سبيرمان براون لتقدير الثبات الداخلي (Internal Consistency) للاختبارات الفرعية، وجد جاكوبس أن القيم تراوحت بين 0.14 للتحليل و0.68 للتبرير الاستنتاجي للنموذج A، في حين كانت 0.42 للتحليل و 0.71 للتبرير الاستنتاجي للنموذج B. وأشار جاكوبس في دراسته إلى الاختلاف في أوساط الفقرة ونمط (Patterns) الارتباطات الداخلية (Intercorrelations) لكلا النموذجين. ويقول الباحث ما زالت هناك عدة تساؤلات تدور حول الاختبار، تتمركز حول الخصائص السيكومترية للنموذجين A ,B. وأوصى مستخدمي الاختبار أن يختبروا بحذر الثبات وارتباط الفقرة مع العلامة الكلية والعلامات الفرعية.

وفي دراسة أجراها لو وثروب (Loo & Thrope, 1999) هدفت إلى البحث في الخصائص السيكومترية للعلامات على اختبار واتسون – جليسر (نموذج S)، وخصوصاً ما يتعلق بصعوبة الفقرة وثبات العلامات بالإضافة إلى تحديد فيما إذا كانت البيانات تدعم البناء النظري لاختبار واتسون– جليسر الذي يقوم على قياس خمس مهارات في أنه يقيس خمسة

مهارات. بدأ بالتعريف باختبار واتسون -جليسر الجديد (نموذج S)، موضحاً أنه اختبار مطور من اختبار واتسون – جليسر القديم (نموذج A)، الشائع الاستعمال والمكون من 80 فقرة. في حين اشتمل اختبار واتسون– جليسر الجديد (نموذج S) على 40 فقرة تم اختيارها من النموذجA . تكونت عينة الدراسة من 271 طالب وطالبة في جامعة في غرب كندا منهم 142 مسجلين في كلية إدارة الأعمال من بينهم 50 طالبة و123 طالب وطالبة في كلية التمريض منهم 117 طالبة. أشارت النتائج إلى أنه لا يوجد فروق بين المتوسطات الحسابية لعلامات طلبة كليتي الإدارة والتمريض على الاختبار الكلي أو أي اختبار فرعي، لكن المتوسطات الحسابية التي ظهرت في الدراسة أقل من المتوسطات الحسابية التي أظهرها الدليل (Manual) الخاص بالاختبار. أما عن الارتباطات الداخلية بين الاختبارات الفرعية الخمسة، فقد تراوحت بين 0.62 إلى 0.73 في الدليل في حين بينت الدراسة أن الارتباطات تراوحت بين 0.53 و 0.64، وقد كان الارتباط لمهارة التقويم في (الاختبار الخامس) هو الأقل ارتباطاً مع العلامة الكلية لكلا العينتين. تراوحت معاملات الصعوبة للاختبار الفرعي الأول بين 0.4 و 0.71 في حين كان معامل الصعوبة للفقرتين 4، 7 أقل من القيمة المقبولة حيث كانت 0.41 و 0.42. أما بالنسبة للاختبارات الفرعية (2– 5) حيث لكل فقرة بديلين، كان معامل الصعوبة للفقرات (24، 30، 31، 36) أكبر من 50% وبذا تبدو أنها فقرات سهلة، في حين كانت معاملات الصعوبة للفقرات (11، 25، 29، 37) أدنى من 50%. وعن أثر الجنس على علامات الاختبارين أظهرت النتائج تحرر العلامات من أثر الجنس، وكذلك من أثر التخصص الأكاديمي، وأوصت الدراسة بإجراء المزيد من البحث في الخصائص السيكومترية للاختبار قبل اعتماده كأداة، مع أن الدراسة بينت أن هذه الأداة تصلح كأداة تعليمية حساسة لأثر البرامج المصممة لتعليم التفكير الناقد.

وفي دراسة اهتمت بدراسة الخصائص السيكومترية للمقاييس، أجرى جاكوبس (Jacobs, 1999) دراسة هدفت إلى البحث في تكافؤ النموذجين A, B لاختبار كاليفورنيا لقياس مهارات التفكير الناقد. تكونت عينة الدراسة من 1461 طالب وطالبة من طلاب السنة الأولى في جامعة ايسترن الخاصة (Eastern Private University)، وتم تقسيم هؤلاء الطلبة عشوائياً إلى مجموعات حيث احتوت كل مجموعة على 25 شخصاً على الأكثر، وتم تطبيق نموذجي الاختبار عليهم عشوائياً. أشارت نتائج الدراسة إلى وجود فروق دالة إحصائياً في صعوبة النموذجين، حيث تبيّن أن النموذج B أكثر صعوبة من النموذج A. وإن أية محاولة لإجراء معادلة (Equating) إحصائية بين النموذجينA, B قد تكون مجرد إضاعة وقت. كما أن أزواج الفقرات لم تكن متكافئة، وأن العلامات (خصوصاً العلامات الفرعية) المعتمدة على بيانات الفقرة لم تكن قابلة للمقارنة بين النموذجين.

وفي دراسة أجراها إلدمان (Eldman, 2002) هدفت إلى دراسة بنية اختبار مينسوتا للتفكير الناقد (Minnesota Test of Critical Thinking) وخصائصه السيكومترية. حيث أشارت الدراسة إلى أن اختبار مينسوتا صمم لقياس المهارات الفرعية الآتية: التفسير والتحليل والتقويم والاستدلال والشرح. تم تطبيق الاختبار على عينة مكونة من 234 طالب وطالبة من طلبة الكليات والجامعات، بالإضافة إلى تطبيق اختبار واتسون – جليسر للتفكير الناقد – (نموذج S) واختبار إنيس واير المقالي. أشارت النتائج إلى أن معامل الثبات كرونباخ ألفا للاختبار كان 0.91، في حين كان معامل الثبات للمهارات الفرعية كما يأتي: مهارة التفسير 0.68، ومهارة التحليل 0.71 والاستدلال 0.66، والتقويم 0.5، والشرح 0.78، وأما فيما يتعلق بالارتباط بين الاختبار المعنى بالدراسة واختبار واتسون – جليسر فقد بلغ 0.66 أما عن ارتباطه مع اختبار إنيس واير فقد كان 0.57. وأوصت الدراسة بأن هذا الاختبار يعد أداة مفيدة لقياس التفكير الناقد لطلبة الكليات والجامعات.

وفي دراسة أخرى قام بها كل من واجنر وهـــارفي (Wagner & Harvey, 2003) هدفت إلى تطوير أداة جديدة لاختبار التفكير الناقد باستخدام نظرية استجابة الفقـــرة، بهـــدف تحسين دالة معلومات الاختبار والخطأ المعياري لأخطاء القياس، وذلك باستخدام أوجه التفكير الناقد لاختبار واتسون – جليسر، ولكن بزيادة عدد البدائل للفقرات، وذلك للتقليل من عامـــل التخمين، توقع الباحثان أن زيادة عدد البدائل في الأداة الجديدة (WAT) سوف يـــؤدي إلـــى تخمين أقل، وتحسّن في دالة معلومات الفقرة والاختبار، وانخفاض الخطأ المعياري في القياس الخاص بواتسون– جليسر، وذلك باستخدام تحليلات نظرية استجابة الفقرة. كما توقع الباحثان أن معامل الثبات الداخلي للاختبار الجديد (WAT) سيكون أعلى من معامل الثبات الـــداخلي لاختبار واتسون –جليسر، وذلك عائد إلى انخفاض في أخطاء القياس. تكوّنت عينة الدراســـة من 407 طالب وطالبة من جامعة (Large Southeastern University) مـــن بيـــنهم 279 طالبة، تتراوح أعمارهم بين 18–27 سنة. تم تطبيق اختبار واتسون – جليسر (النموذج A) بالإضافة إلى اختبار (WAT) على نفس العينة. أشارت نتائج التحليل إلى ما يأتي: في حالـــة وجود البديلين في اختبار واتسون – جليسر لأربعة اختبارات فرعية من بين خمسة، كانت قيم المعالم (C) جوهرية وذات دلالة. وهناك تحسّن في دقة القياس خصوصاً بـــين المفحوصـــين متدنيي القدرة في الأداة الجديدة، وكانت أخطاء القياس في اختبار واتسون– جليسر أكبر مـــن أخطاء القياس للاختبار الجديد (WAT) بمقدار 50% على الأقل، في حالة θ تحت الـــصفر، وفوق الضعف في حالة (θ تحت -1) وحتى عندما تكـــون (θ= 1) فـــإن أخطـــاء القيـــاس لواتسون–جليسر غالباً ما تكون ضعف أخطاء القياس للاختبار الجديد (WAT).

وفي الأردن أجرى شطناوي (2003) دراسة هدفت إلى تقنين اختبار كاليفورنيا لمهارات التفكير الناقد واشتقاق معايير أداء طلبة الجامعات الأردنية عليه بعد تطويره وتعديله للبيئة الأردنية. تكونت عينة التقنين من 1485 طالب وطالبة من جامعة مؤتة موزعين على كافة الكليات العلمية والإنسانية. أظهرت نتائج الدراسة فروقاً دالة إحصائياً في متوسطات أداء الطلبة تعزى للتخصص الدراسي والعمر، والتفاعل بين التخصص الدراسي والجنس والعمر. ولم تشر النتائج إلى وجود فروق ذات دلالة إحصائية في متوسطات أداء الطلبة تعزى لمتغير الجنس أو العمر. وقد تم تحديد المجموعات المعيارية واشتقاق الرتب المعيارية والعلامات التائية لأداء طلبة الكليات العلمية، وطلبة الكليات الإنسانية وطلبة الكليات مجتمعة على العلامات الكلية والمهارات الفرعية للاختبار.

وفي دراسة أجريت في جامعة تينسي التكنولوجية (Tennessee Technological University) بهدف البحث عن أداة جديدة لتقييم مهارات التفكير الناقد، قام الفريق المكون من ستين، وهاينس وانترستين (Stein, Haynes & Unterstein, 2003) بتحديد المهارات الهامة والضرورية وتجريبها في كل مرة والعمل على التعديل ووضع معايير التصحيح. ولتحقيق صدق المحك تمت مقارنة الأداء على هذا الاختبار مع مقاييس أخرى كالتحصيل الأكاديمي، بالإضافة إلى مقارنة نتائج هذه الأداة مع نتائج اختبار كاليفورنيا، بينت نتائج المقارنات ارتباطات متوسطة لكنها ليست عالية. طبقت الأداة على عينة عشوائية طبقية تم اختيارها من الطلبة المتوقع تخرجهم من أربع كليات (آداب وعلوم، تربية، هندسة، إدارة أعمال). ولتقويم الأداة الجديدة فقد تم تطبيق اختبار كاليفورنيا للتفكير الناقد على 119 طالباً، بالمقابل تم تطبيق الأداة المصممة (CAT) على 104 طلاب، من بين هؤلاء جميعاً 64 طالباً طبّق عليهم الاختباران. استخدم تحليل التباين المصاحب ANCOVA لإزالة الأثر بين

المجموعات، ولتقويم حساسية الاختبار وفحص التحسن الممكن في مهارات التفكير الناقد، تمت مقارنة نتائج اختبار القبول لسنة أولى مع نتائج السنة الأخيرة في الجامعة، ولقد أظهرت النتائج زيادة ذات دلالة في علامات اختبار التفكير الناقد عند انتقال الطلبة من السنة الأولى إلى السنة الأخيرة. وقد تم تطبيق الاختبار الجديد (CAT) بهدف: معرفة فيما إذا كان الاختبار حساساً بشكل كافٍ لقياس الاكتساب الذي يمكن تحقيقه في مساق واحد صمم لتحسين التفكير الناقد. وبفرض أن المساق لا يعمل على تحسين مهارات التفكير الناقد يريد الباحثون معرفة كيف يكون مقدار الثبات (Reliable) للاختبار عند أخذه مرتين وبوقتين مختلفين. وإن تقويم أثار المساق المصمم لتحسين مهارات التفكير الناقد وحل المشكلة على أداء الاختبار. ولتحقيق الأهداف المذكورة سابقاً، تم تطبيق الأداة في مساقين مختلفين في العلوم الاجتماعية، حيث صمم أحد المساقين بشكل خاص لتحسين مهارات التفكير الناقد وحل المشكلات. أما المساق الآخر فقد خصص للمجموعة الضابطة. وتم إخضاع الطلبة في كلا المساقين لاختبار قبلي وآخر بعدي في التفكير الناقد بفاصل زمني تقريباً 13 أسبوع، كان هناك 16 طالباً في المجموعة الضابطة أخذوا الاختبارين القبلي والبعدي حيث بينت النتائج أنه لا فروق ذات دلالة بين أداء الطلبة على الاختبارين. وأن هناك 19 طالباً في المجموعة التجريبية أخذوا الاختبارين القبلي والبعدي، وبينت النتائج وجود فروق ذات دلالة بين الاختبارين ولصالح الاختبار البعدي. وأشارت النتائج إلى أن اختبار (CAT) قد يكون حساساً إلى حدٍ كافٍ بما فيه الكفاية لاكتشاف الآثار الإيجابية للمساق على التفكير الناقد لدى الفرد. وأيدت نتائج الدراسة أن الاختبار قادر على التعريف بالمساقات وطرق التدريس التي تعزز وترفع مستوى التفكير الناقد لدى الطلبة.

وفي دراسة أخرى أجراها كل من جادزيلا وستيكس وماستن (Gadzella, Stacks & Masten, 2005) هدفت إلى البحث فيما إذا كان اختبار واتسون – جليسر (النمــوذج S) أداة صادقة وثابتة لقياس مهارات التفكير الناقد. تكونت عينة الدراسة من 137 طالب وطالبة مسجلين في مساق علم النفس التربوي (مساق أساسي لبرنامج تدريب المعلمين) فـــي جامعـــة الولايــــة (State University)، من بينهم 109 طالبة. تم تحليل علامات الطلبة في العينة على كل من اختبار واتسون – جليسر (النموذج S) وعلامات الطلبة أيضاً في مـساق علـم النفس التربوي التي أخذت من الفقـرات الموضـوعية مـن (Assessment Package for Woolfolk's Educational Psychology) والتي تقيس الاستيعاب، والتفـسير، وتـذكر الحقائق وفي أربعة ميادين هي: طرق البحث، التطور الاجتماعي والأخلاقي، التطور المعرفي وتعلم السلوك والقياس والتقويم. تراوحت علامات الطلبة على اختبار واتسون– جليسر بـين 14 و 39 بمتوسط حسابي 24.2 وانحراف معياري 5، علماً بأن العلامة الكلية على الاختبار 40، وتراوحت علامات المساق من 43% إلى 99%. وتم حساب معامل الثبات لكل اختبـار فرعي وللاختبار الكلي لاختبار واتسون– جليسر، حيث كانت قيمـة Kr-20 تـساوي 0.76 للمجموعة الكلية، ومعامل الثبات النصفي للفقرات الفردية (r = 0.44)، وهو أقل من المذكور في دليل اختبار واتسون –جليسر. أما صدق الأداة فقد تـم تحديـده عـن طريـق ارتبـاط الاختبارات الفرعية والكلية مع علامات المساق، حيث كان معامل الارتباط بين العلامة الكلية لاختبار واتسون –جليسـر وعلامة المساق (r = 0.31)، وهو ضمن مدى العلامات المذكور في الدليل، ولأن مدى علامات المساق كبير، فقد تم تقسيم المجموعة إلى قسمين: القسم الأول: مجموعة العلامات العالية وهم الطلبة الذين حصلوا على العلامـات (A, B) وعـددهم 113 طالباً في حين كان عدد طلبة القسم الثاني 24 طالباً وهم الذين حصلوا على العلامات (C) أو

أقل. أشارت نتائج تحليل العلامات إلى أن علامات التفكير الناقد كانت أعلى للطلبة ذوي العلامات العالية في مهارتي الاستنتاج وتقويم الحجج والاختبار الكلي بشكل عام. أشار الباحثون في هذه الدراسة أن اختبار واتسون – جليسر (S)، أداة جيدة لقياس مهارات التفكير الناقد.

ويمكن إجمال نتائج الدراسات السابقة على النحو الآتي:

- لا تخضع المقاييس المتوفرة حالياً إلى معايير القياس الموضوعي حيث تتأثر خصائصها السيكومترية من صدق وثبات بخصائص مجموعة من الأفراد التي تخضع لها، ومع ذلك لازلت هذه المقاييس مثل اختبار واتسون– جليسر، كاليفورنيا، كورنل، تستخدم على نطاق واسع.
- رغم قدم اختبار واتسون – جليسر إلا أنه ما زال الأكثر استخداماً (Wagner & Harvey, 2003).
- اعتماد بعض الباحثين على بناء أدوات خاصة بدراساتهم (السيد، 1995؛ Stein, Haynes & Unterstein, 2003).
- استخدام نظرية استجابة الفقرة على نطاق ضيق في بناء أدوات قياس التفكير الناقد. وهذا ربما عائد لحداثة الاهتمام بالنظرية الحديثة في القياس ولحداثة أيضاً الدراسات التي أفرزت سلبيات المقاييس الموجودة حالياً وللتأخر في ظهور تعريف شامل للتفكير الناقد.

ومن هذا المنطلق تنشأ الحاجة إلى بناء وتطوير اختبار للتفكير الناقد يتمتع بخصائص القياس الموضوعي.

وقد جاءت هذه الدراسة تلبية لهذه الحاجة، وتحاول الباحثة من خلالها بناء اختبار لقياس مهارات التفكير الناقد وفق نظرية استجابة الفقرة.راعت الباحثة في هذه الدراسة أن يشتمل الاختبار الواسع من مهارات التفكير الناقد من خلال:

- اعتماد التعريف الأكثر شمولية لمفهوم التفكير الناقد والذي تقريباً اتضح أكثر في السنوات الأخيرة كما يرى بول ونوسج (Paul & Nosich, 1992).
- إدخال مهارات جديدة للاختبار، لم تكن موجودة في الاختبارات السابقة تنسجم ومتطلبات العصر الحالي مثل مهارة التبرير الإحصائي والاحتمالي.
- اشتمال الاختبار على مواقف حياتية حقيقية تضع المفحوص أمام موقف يعيشه وهذا الموقف يحوي مشكلة يحتاج لحلها إحدى مهارات التفكير الناقد على الأقل.
- الاعتماد على تجمع كبير من الفقرات (Pool Items) في بناء الاختبار.
- بناء الاختبار وفق نماذج نظرية استجابة الفقرة.

الفصل الثالث

الطريقة والإجراءات

يتناول هذا الفصل وصفاً لأفراد الدراسة، وطريقة اختيارها، ووصفاً لطريقة بناء الاختبار (أداة الدراسة) وإجراءات تطبيقه، كما يتناول وصفاً مفصلاً لخطوات إجراءات الدراسة والمعالجة الإحصائية التي أُجريت.

أفراد الدراسة:

تكونت عينة الدراسة من 1188 طالباً وطالبة، موزعين على أربع فئات هي: طلبة المرحلة الأساسية العليا، وطلبة المرحلة الثانوية، وطلبة جامعيين، ويبيّن الجدول 1 توزيع أفراد عينة الدراسة وفق متغيري الجنس والمستوى الأكاديمي.

جدول 1

توزيع أفراد عينة الدراسة الكلية وفق الجنس والمستوى الأكاديمي

المستوى الأكاديمي	ذكور	إناث
طالب في المرحلة الأساسية العليا	192	266
طالب في المرحلة الثانوية	302	294
طالب جامعي	76	58

وفيما يأتي توضيح لخصائص عينة كل مستوى أكاديمي.

1. طلبة المرحلة الأساسية العليا:

تم اختيار سبع مدارس عشوائياً من بين المدارس في مديرية التربية والتعليم لمنطقة اربد الأولى للعام الدراسي 2005-2006م التي فيها شعب للصف التاسع أو الصف العاشر. وتكونت المدارس المختارة من مدرستين للذكور وأربع مدارس للإناث، ومدرسة مختلطة هي مدرسة الملك عبد الله الثاني للتميّز. ومن كل مدرسة اختيرت شعبة أو شعبتان عشوائياً من شعب الصف التاسع أو الصف العاشر، ويبيّن الجدول 2 أسماء المدارس، وعدد الشعب، وعدد الطلبة الذين اختيروا للعينة من الجنسين.

جدول 2
توزيع أفراد العينة من المرحلة الأساسية العليا وفق الجنس ومستوى الصف

		الصف			
		التاسع		العاشر	
	اسم المدرسة	عدد الشعب	عدد الطلبة	عدد الشعب	عدد الطلبة
ذكور	حمزة بن عبد المطلب الأساسية	1	44	1	44
	الملك عبد الله الثاني للتميّز	-	-	1	19
	المثنى بن حارثة الأساسية	1	43	1	42
إناث	رقية بنت الرسول الأساسية	2	94	1	44
	طبريا الثانوية	-	-	1	42
	النهضة الأساسية	1	15	1	15
	بشرى الثانوية	-	-	1	36
	الملك عبد الله الثاني للتميّز	-	-	1	20

2. طلبة المدارس الثانوية:

تم اختيار ست مدارس عشوائياً من بين المدارس التي فيها شعب للصف الأول الثانوي بمسارين أكاديميين على الأقل في مديرية التربية والتعليم لمنطقة اربد الأولى للعام الدراسي 2005-2006م، وتضمنت المدارس المختارة ثلاث مدارس للذكور، وثلاث مدارس للإناث. ومن كل مدرسة اختيرت شعبة أو شعبتان عشوائياً، من شعب الصف الأول الثانوي. ويبين الجدول 3 أسماء المدارس وعدد الشعب، وعدد الطلبة الذين اختيروا للعينة موزعين حسب الجنس والمسار الأكاديمي.

جدول 3

توزيع أفراد العينة من الصف الأول الثانوي وفق الجنس والمسار الأكاديمي

		المسار الأكاديمي					
		علمي		أدبي		إدارة معلوماتية	
	اسم المدرسة	عدد الشعب	عدد الطلبة	عدد الشعب	عدد الطلبة	عدد الشعب	عدد الطلبة
ذكور	الوليد بن عبد الملك الثانوية	2	58	1	24	–	–
	علي خلقي الشرايري الثانوية	2	70	–	–	–	–
	رنوجي لثانوية	2	75	1	40	1	35
إناث	نور الحسين الثانوية	1	31	1	30	2	78
	طبريا الثانوية	1	42	1	38	–	–
	بشرى الثانوية	2	50	1	25	–	–

3. الطلبة الجامعيون:

تم اختيار الطلبة الجامعيين بشكل قصدي من جامعتي: اليرموك والعلوم والتكنولوجيا الأردنية، وذلك بسبب سهولة وصول الباحثة إلى هاتين الجامعتين، ولقناعة الباحثة أيضاً بأن الطلبة في الجامعات الحكومية يخضعون لنفس شروط القبول. تم اختيار شعبتين من كلية العلوم وشعبتين من كلية الآداب بشكل عشوائي من جامعة اليرموك. بالإضافة إلى اختيار كلية الطب البشري في جامعة العلوم والتكنولوجيا بشكل قصدي لأن جميع طلبة هذه الكلية يعدوا من ذوي التحصيل العالي لاعتماد معدل القبول 85% وهو أعلى معدل قبول في جميع التخصصات في الجامعات الأردنية، وتم اختيار طلبة السنتين الرابعة والسادسة بشكل عشوائي، تم بعد ذلك اختيار شعبة من طلبة السنة الرابعة وشعبة من طلبة السنة السادسة بشكل عشوائي. يوضح الجدول 4 توزيع أفراد العينة وفق التخصص وعدد الطلبة الذين تم اختيارهم من الجنسين.

جدول 4

توزيع أفراد عينة الدراسة من الطلبة الجامعيين وفق الجنس والتخصص

التخصص	ذكور	إناث
علوم	30	17
آداب	28	35
طب بشري	18	6

بناء فقرات الاختبار:

روعي في بناء الاختبار ما يأتي:

- تحديد تعريف التفكير الناقد، وحصر المهارات التي تندرج تحت هذا التعريف. حيث قامت الباحثة وبعد الاطلاع على تعريفات مختلفة للتفكير الناقد بتعريف التفكير الناقد بصورته الشمولية بالاعتماد على تعريف كاليفورنيا (Facione, & Facione, 1998) وهالبرن (Halpern, 2001)؛ وإنيس (Ennis, 1993) للتفكير الناقد على أنه أسلوب في التفكير المنظم والمركّز والمبرر بغرض حل المسألة وصياغة الاستدلالات وحساب الاحتمالات وعمل القرارات. وتندرج تحت هذا التعريف المهارات الأساسية الآتية: التحليل وتتضمن مهارتين فرعيتين، والتفسير وتتضمن ثلاث مهارات فرعية، والتقويم وتتضمن ثلاث مهارات فرعية، والشرح وتتضمن ثلاث مهارات فرعية، والاستدلال وتتضمن أربع مهارات فرعية، والتبرير الإحصائي والاحتمالي وتتضمن مهارتين فرعيتين، والتعرف إلى الافتراضات المنصوص وغير المنصوص عليها، والنظر إلى الأشياء بعمق وتوسع (التركيز) ويتضمن مهارتين فرعيتين.
- بناء تجمع من الفقرات (Item Pool) يتكون من 228 فقرة من نوع الاختيار من متعدد بأربعة بدائل لكل فقرة منها بديل واحد فقط صحيح، يقيس الاختبار (Item Pool) المهارات الآتية: التحليل وخصص لها 26 فقرة، والتفسير وخصص لها 27 فقرة، والتقويم وخصص لها 26 فقرة، والشرح وخصص لها 22 فقرة، والاستدلال وخصص لها 57 فقرة، والتبرير الإحصائي والاحتمالي وخصص لها 27 فقرة، والتعرف إلى الافتراضات المنصوص وغير المنصوص عليها وخصص لها 25 فقرة، والنظر إلى الأشياء بعمق وتوسع (التركيز) وخصص لها 18 فقرة.

- عرض التعريف والمهارات الأساسية والفرعية مع الفقرات على هيئة تحكيم (ملحق 1) تكونت من 18 أستاذ جامعي من حملة درجة الدكتوراه في تخصصات مختلفة، بالإضافة إلى أحد أعضاء هيئة التدريس في الجامعة والذي يحمل درجة الماجستير. كما شملت هيئة التحكيم سبعة مشرفين وموظفي تربية من حملة درجتي الدكتوراه أو الماجستير في تخصصات مختلفة، والذين خضعوا لدورات في التفكير الناقد عقدتها وزارة التربية والتعليم. وضمت هيئة التحكيم أيضاً معلمة تحمل درجة البكالوريوس في الفيزياء وخضعت أيضاً لدورة في التفكير الناقد عقدتها وزارة التربية والتعليم في أثناء فترة إعداد الأطروحة، وطلب من كل محكم بيان رأيه في الفقرات والمهارات الأساسية والفرعية من حيث:

- انسجام المهارات الأساسية مع التعريف.
- انسجام المهارات الفرعية مع المهارة الأساسية.
- وضوح الفقرات لغوياً.
- تمثيل الفقرة للمهارة التي تندرج تحتها.
- ملائمة بدائل الفقرة الواحدة.

وتم تعديل في الفقرات والمهارات الأساسية والفرعية في ضوء مقترحاتهم.

كما استخدمت الباحثة ما يُعرف بالمجموعات المركزة (Focus Groups)؛ إذ استفادت من مجموعات المعلمين في أثناء الدورات بحكم عملها كمشرفة في وزارة التربية والتعليم في مناقشة فقرات الاختبار والحكم على مدى اتساق الفقرات وسلامتها اللغوية، حيث قامت بتعديل نصوص أو إلغاء بعض الفقرات، وإضافة فقرات أخرى بالاعتماد على أفكارهم وآرائهم وتساؤلاتهم. كما لجأت إلى مجموعات مركّزة من مجموعة من طلبة المدارس والجامعات كانت الباحثة تجلس معهم وتناقشهم في بعض الفقرات وأيضاً تطلب منهم التحدث حول الفقرة ومناقشتها. وتستمع إلى الأسئلة التي تُطرح حول الفقرة، والنقد الموجه للفقرة.

ومن المجموعات الأخرى التي شكلتها الباحثة مجموعة من الراشدين من ذوي الاختصاصات المختلفة كالأطباء والمهندسين والمحامين الذين أثروا الاختبار بآرائهم وأفكارهم، ومجموعة من المشرفين الذين يعملون مع الباحثة بنفس المديرية الذين ناقشوا الفقرات في جلسات متعددة، من حيث صياغتها وملاءمتها ومدى اتساق البدائل المطروحة.

وبعد انتهاء عملية التحكيم المتشعبة والممتدة في الفترة الواقعة ما بين 2005/6/16 إلى 2005/8/22 والأخذ بملاحظات المحكمين، استقرت الأداة على 195 فقرة.

التجريب الأوليّ للاختبار:

نظراً لطول الاختبار وصعوبة تطبيقه بصورته الكلية (195 فقرة) لنفس الأفراد حيث يحتاج إلى ما يزيد عن أربع ساعات لتطبيقه، كان لابد من تقسيمه إلى عدة نماذج، ولتسهيل عملية مقارنة العلامات على هذه النماذج بصورة عادلة كان لابد من أن تكون هناك فقرات مشتركة بين هذه النماذج. وقد روعي في الفقرات المشتركة أن تغطي المهارات الأساسية التي يغطيها الاختبار بصورته الكلية. وبناءً على ذلك قامت الباحثة بتقسيم الاختبار الكلي (195 فقرة) إلى خمسة نماذج اختبارية بواقع 51 فقرة لكل نموذج من بينها 15 فقرة مشتركة (Common Items). حيث راعت الباحثة في أثناء عملية التقسيم توزيع الفقرات التي تندرج تحت المهارة الواحدة على النماذج الخمسة بالتساوي والاحتفاظ بباقي القسمة كفقرات مشتركة، مع مراعاة شمول الفقرات المشتركة كافة المهارات الأساسية حيث احتوى كل نموذج 15 فقرة مشتركة تقيس المهارات الآتية: التحليل ولها فقرة واحدة، والتقويم ولها ثلاث فقرات، والشرح ولها فقرتان، والاستدلال ولها أربع فقرات، والتبرير الإحصائي والاحتمالي ولها ثلاث فقرات، والتعرّف على الافتراضات المنصوص وغير المنصوص عليها ولها فقرتان. وأما بخصوص الفقرات غير المشتركة وعددها 36 فقرة لكل نموذج، فقد خُصصت فقرتان لمهارة التحليل

وست فقرات لمهارة التفسير، وسبع فقرات لمهارة التقويم، وفقرتان لمهارة الشرح و13 فقرة لمهارة الاستدلال وثلاث فقرات لمهارة التبرير الإحصائي والاحتمالي، وفقرة واحدة لمهارة التعرّف على الافتراضات المنصوص وغير المنصوص عليها ومهارتان للتركيز. وبعد ذلك قامت بتطبيق الاختبار بنماذجه الخمسة على مجموعة من الطلبة في مدينة جرش بلغ عددهم 156 طالب وطالبة، حيث قام كل طالب بالإجابة عن نموذج واحد فقط، من النماذج الخمسة، ويبيّن الجدول 5 توزيع أفراد العينة الاستطلاعية وفق الجنس ومستوى الصف.

جدول 5

توزيع أفراد العينة الاستطلاعية وفق الجنس ومستوى الصف

الصف	ذكور		إناث	
	عدد الشعب	عدد الطلبة	عدد الشعب	عدد الطلبة
أول ثانوي علمي	–	–	2	54
أول ثانوي أدبي	–	–	1	26
عاشر أساسي	2	76	–	–

حيث انحصر الغرض من التجريب على:

- تحديد الزمن الذي يستغرقه الطلبة في الإجابة عن النماذج الاختبارية.
- جمع أي ملحوظات حول الفقرات والكشف عن الفقرات الغامضة أو العبارات التي كثر التساؤل حولها.
- التعرف على مستوى صعوبة الفقرات وقدرتها التمييزية.
- دراسة فاعلية المموهات.
- دراسة الخصائص السيكومترية للفقرات من أجل تحديد الفقرات المناسبة منها.

الصورة النهائية للاختبار:

بعد دراسة إجابات العينة الاستطلاعية أثناء عملية التجريب، ودراسة معاملات التمييز للفقرات تم حذف الفقرات ذات التمييز المنخفض (أقل من 0.2) وبلغ عدد الفقرات المحذوفة 106 فقرات، وتكوّن الاختبار في صورته النهائية من 89 فقرة، كانت معاملات الصعوبة لها تتراوح بين 0.21 و0.70 ومعاملات التمييز تتراوح بين 0.20 و0.68. ويشير (الملحق 3) إلى عينة من هذه الفقرات.

ونظراً لأن الاختبار بصورته النهائية (89 فقرة) يحتاج إلى ساعتين لتطبيقه، فقد ارتأت الباحثة أن تطبقه في جلستين، ومن أجل ذلك ولتسهيل عملية التطبيق لهذه الفقرات فقد تم تقسيمها عشوائياً إلى قسمين: يتكون القسم الأول منها من 45 فقرة، والثاني من 44 فقرة.

تطبيق الاختبار:

مرت عملية التطبيق بالمراحل الآتية:

- التنسيق المسبق مع مديري ومديرات المدارس للاتفاق على الموعد الذي يطبق فيه الامتحان وآلية التطبيق، كذلك الالتقاء بالمعلمين الذين سيطبّق الاختبار في حصصهم لتوضيح تعليمات الاختبار لهم وآلية التطبيق.
- الالتقاء مع الطلبة في اليوم المحدد لإجراء الاختبار قبل موعد الاختبار والتحدث إليهم حول الاختبار بإعطاء فكرة عن التفكير الناقد وتعريفه، وأهميته في حياة الطالب بشكل خاص والإنسان بشكل عام، وحث الطلبة على توخي الدقة في أثناء الإجابة مع التوصية بالإجابة عن جميع فقرات الاختبار، كما تم التوضيح للطلبة كيفية الإجابة بتظليل الدائرة التي تندرج تحت رمز الإجابة التي يختارها الطالب مع إعطاء مثال مع التعليمات التي تمت قراءتها للطلبة.

- قامت الباحثة بتوزيع كراسات الاختبار للطلبة ونماذج الإجابة بحيث تم توزيع القسمين الأول والثاني بنفس الجلسة حيث إن الطالب الذي يجيب عن اختبار القسم الأول يجيب الطالب المجاور له عن اختبار القسم الثاني وهكذا، وذلك للحد من عملية الغش والكلام بين الطلبة، ولضمان توفير نفس ظروف الامتحان لقسمي الاختبار.
- وبعد انتهاء الجلسة الأولى سُمح للطلبة باستراحة لمدة عشر دقائق ليعودوا بعدها للجلوس من أجل الإجابة عن فقرات القسم الآخر من الاختبار وبنفس الآلية.
- جمعت أوراق الإجابة وكراسات الاختبار بعد ذلك من الجميع وتم التنسيق بين إجابة القسم الأول، والثاني لكل طالب وذلك عن طريق مطابقة الاسم والرقم المتسلسل الموجود على ورقة إجابة الطالب، وتم استبعاد الاستجابات غير الجدية أو ذات النمطية الواحدة في الإجابة.

أما بالنسبة للتطبيق في الجامعات فعن طريق الاتفاق المسبق مع الدكتور مدرس المساق تم الاتفاق على الموعد الذي سيطبق فيه الامتحان حيث كان التطبيق على النحو الآتي: في المحاضرة الأولى ومدتها ساعة يأخذ جميع طلبة الشعبة نفس الجزء (من الاختبار) وبعد انتهاء المحاضرة تجمع كراسات الاختبار وأوراق الإجابة، وفي المحاضرة التي تليها مباشرة في اليوم التالي يأخذ جميع طلبة الشعبة الجزء الثاني من الاختبار.

وقد تم الاستغناء أيضاً عن الاستجابات غير الجدية أو أوراق الإجابة التي تغيب فيها الطالب عن أحد جزأي الاختبار. هذا وقد طبق الاختبار في الفترة الواقعة ما بين 10/12- 2005/10/23 لطلبة المدارس وفق الجدول 6.

جدول 6

أسماء المدارس والأيام التي تم تطبيق الدراسة فيها

المدرسة	الصف	اليوم والتاريخ
رقية بنت الرسول الأساسية للبنات	تاسع، عاشر	أربعاء 10/12
نور الحسين الثانوية للبنات	أول علمي، أول معلوماتية	أربعاء 10/12
طبريا الثانوية للبنات	عاشر، أول علمي، أول أدبي	خميس 10/13
حمزة بن عبد المطلب الأساسية للبنين	تاسع، عاشر	أحد 10/16
الوليد بن عبد الملك الثانوية للبنين	أول علمي، أول أدبي	اثنين 10/17
النهضة الأساسية الخاصة للبنات	عاشر، تاسع	اثنين10/17
بشرى الثانوية للبنات	عاشر، أول علمي، أول أدبي	ثلاثاء 10/18
علي خلقي الثانوية للبنين	أول علمي	أربعاء 10/19
مدارس الملك عبد الله الثاني للتميّز	عاشر، ذكور وإناث	أربعاء 10/19
الزرنوجي الثانوية الشاملة للبنين	أول علمي، أول معلوماتية	الخميس 10/20
المثنى الأساسية للبنين	تاسع، عاشر	أحد 10/23

ولطلبة جامعة اليرموك في الفترة من 2005/11/8 إلى 2005/11/17 وفق الجدول 7.

جدول 7

توزيع شعب طلبة جامعة اليرموك على الأيام التي تم تطبيق الاختبار فيها

التخصص /الشعبة	اليوم/ التاريخ
الآداب (1)	ثلاثاء 11/8، ثلاثاء 11/15
الآداب (2)	ثلاثاء 11/15، خميس 11/17
العلوم (1)	ثلاثاء 11/8، ثلاثاء 11/15
العلوم (2)	ثلاثاء 11/15، خميس 11/17

أما بالنسبة لطلبة كلية الطب في جامعة العلوم والتكنولوجيا، فقد تم تطبيق الاختبار أثناء تواجدهم للتطبيق العملي في مستشفى الأميرة رحمة في يومين وفق الجدول 8 الآتي:

جدول 8

توزيع طلبة كلية الطب على الأيام التي تم تطبيق الاختبار فيها

المستوى	اليوم والتاريخ
سنة رابعة	أربعاء 11/16 ، خميس 11/17
سنة سادسة	أربعاء 11/16 ، خميس 11/17

المعالجة الإحصائية:

تم إدخال البيانات إلى ذاكرة الحاسوب بعد أن جمعت أوراق الإجابة واستبعدت الاستجابات النمطية والاستجابات غير التامة. تم إعطاء الطالب الرقم 1 إن اختار البديل (أ) والرقم 2 إن اختار البديل (ب) والرقم 3 إن اختار البديل (جـ) والرقم 4 إن اختار البديل (د)، كما تم إعطاء الطالب الرقم 5 إن ترك الفقرة دون إجابة حيث اعتبر هذا الرقم إجابة خاطئة عند إجراء التحليل.

كما تم إدخال مفتاح التصحيح في ملف خاص به ليتم التصحيح بشكل آلي، حيث أعطي الأمر للبرنامج إعطاء العلامة واحد للإجابة الصحيحة، والعلامة صفر للإجابة الخاطئة. وبهذا حُولت جميع البيانات بعد عملية التصحيح الآلي إلى صفر وواحد.

تم إجراء المعالجات الإحصائية الآتية لتحقيق أغراض الدراسة:

أولاً: مؤشرات الصدق:

تم حساب المؤشرات الآتية للتأكد من صدق الاختبار:

أ. صدق المحتوى Content Validity:

تم التحقق من صدق المحتوى من خلال عرض التعريف الذي تبنته الباحثة، والمهارات التي رأت أنها تتسق مع التعريف بالإضافة إلى فقرات الاختبار على محكمين من

ذوي الخبرة والاختصاص. وبعد ذلك تم الأخذ برأي المحكمين، حيث أخذت الفقرات التي أجمع المحكمون على ملاءمتها في حين تم حذف أي فقرة أجمع ثلاثة محكمين على الأقل على عدم ملاءمتها. كذلك تم حذف كل مهارة أو مهارة فرعية أشار على عدم ملاءمتها ثلاثة محكمين على الأقل من مجموعة المحكمين. واهتمت الباحثة بالاقتراحات والملاحظات الإضافية التي قدمها المحكمون. إذ إنهم قدموا اقتراحات قيّمة ساعدت الباحثة في إعادة النظر في بعض المهارات ومراجعة عدد من الفقرات التي كان يعوزها الدقة أو الوضوح في الصياغة.

ب. صدق البناء: (Construct Validity):

تم التحقق من صدق البناء عن طريق مقارنة الأوساط الحسابية للعلامات على الاختبار للفئات الطلابية الآتية:

- طلبة كلية الطب مع عينة عشوائية من طلبة جامعة اليرموك تماثلها في العدد.
- ذوي التحصيل الدراسي المتدني والعالي وذلك من خلال مقارنة أداء عينة من طلاب وطالبات الصف العاشر من ذوي التحصيل المتدني الذين حققوا معدلات مدرسية 65 فما دون تم اختيارهم من عينة الدراسة مع عينة من طلبة الصف العاشر في مدارس الملك عبد الله الثاني للتميّز والذين تم اختيارهم لدخول هذه المدرسة بناء على معايير منها التفوق في التحصيل الدراسي.

2. استخراج دلالة الفروق في الأداء على الاختبار بين الجنسين في مستويات تعليمية مختلفة (أساسي عالي ممثلاً بطلبة الصفين التاسع والعاشر، ثانوي ممثلا بطلبة الصف الأول الثانوي، وجامعي) وذلك باستخدام التصميم العاملي 2 × 3. واستخدام اختبارات بعدية لتحديد الفروق بين المتوسطات الحسابية الدالة إحصائياً.

ج. صدق المحك: (Criterion Validity):

تم اختيار مدرستين بشكل عشوائي من المدارس الثانوية التابعة لعينة الدراسة واحدة للذكور وأخرى للإناث، وتم اختيار شعبة من شعب الصف الأول الثانوي من شعب الدراسة بشكل عشوائي من كل مدرسة، بلغ عدد الطلبة فيهما 82 طالب وطالبة من بينهم 42 طالبة. طبق على هذه المجموعة اختبار كاليفورنيا 2000 المعرب والمعد للبيئة الأردنية. وتم حساب معامل ارتباط بيرسون بين علامات الطلبة على الاختبار الذي أعدته الباحثة واختبار كاليفورنيا 2000.

قامت الباحثة بعد ذلك بحساب معامل الارتباط بين علامات نفس العينة السابقة على الاختبار الذي أعدته الباحثة ومعدل علاماتهم المدرسية.

ثانياً: مؤشرات الثبات:

باستخدام برمجيةBilog-Mg تم تقدير ثبات الاختبار النظري (Theoretical Reliability) والثبات الإمبريقي (التجريبي) (Empirical Reliability) للاختبار بصورته النهائية وفق النماذج الثلاثة.

كما تم تقدير ثبات الاتساق الداخلي (كودر ريتشاردسون 20 (KR-20)) من خلال برمجية SPSS للاختبار بصورته النهائية وفق النماذج الثلاثة.

ثالثاً: التحقق من افتراضات نماذج نظرية استجابة الفقرة:

- أحادية البعد (Unidimensionality):

تم التحقق من هذا الافتراض بعدة مؤشرات:

أ. المؤشرات التي اعتمدت على التحليل العاملي للمكونات الأساسية (Principal Components):

تم دراسة المؤشرات الآتية وذلك من خلال برمجية SPSS:

1. نسبة التباين المفسر للعامل الأول.
2. نسبة الفرق بين الجذر الكامن الأول والجذر الكامن الثاني إلى الفرق بين الجذر الكامن الثاني والجذر الكامن الثالث.
3. نسبة الجذر الكامن الأول إلى الجذر الكامن الثاني. ودراسة العلاقة بين الجذر الكامن الثاني وبقية الجذور التي تليه بالإضافة إلى رسم بياني للجذور الكامنة للعوامل المكونة للاختبار (Scree Plot).

ب. المؤشرات التي اعتمدت على الثبات:

1. تم حساب معامل الارتباط بين الأداء على الفقرة والأداء على الاختبار بعد حذف الفقرة (corrected item Total Correlation).
2. معامل الثبات (KR-20).

- **الاستقلال الموضعي (Local Independence):**

يعد تحقق افتراض الاستقلال الموضعي نتيجة محصلة من تحقق افتراض أُحادية البعد (Hulin, Drasgow & Parsons, 1983; deGruijter& Van der Kamp, 2005). في حين ينظر إليه بعض الباحثين على أنه افتراض مكافئ لافتراض أُحادية البعد (Hambleton, 1989; Hambleton & Swaminathan, 1985)، فلذلك فإذا ما تحققت الباحثة من تحقق افتراض أُحادية البعد فإن الاستقلال الموضعي يتحقق أيضاً.

رابعاً: مطابقة الأفراد للنماذج الثلاثة في نظرية استجابة الفقرة:

تم من خلال برمجية Bilog-Mg مطابقة الأفراد للنماذج الثلاثة وباستخدام اختبار كاي تربيع عند مستوى الدلالة (α = 0.01).

خامساً: مطابقة الفقرات للنماذج الثلاثة في نظرية استجابة الفقرة:

تم استخدام برمجيةBilog-Mg لمطابقة الفقرات على النماذج اللوجستية الثلاثة لنظرية استجابة الفقرة من خلال استخدام اختبار كاي تربيع عند مستوى الدلالة ($\alpha = 0.01$).

سادساً: تقدير معالم الفقرات:

تم استخدام برمجيةBilog-Mg لتقدير معالم الفقرة بطريقة الأرجحية العظمى الهامشية ((Marginal Maximum Likelihood (MML)).

سابعاً: تقدير معالم القدرة:

تم تقدير معلَم القدرة θ عن طريق مقدرات الأرجحية العظمي (Maximum Likelihood Estimation (ML)).

ثامناً: حساب دالة معلومات الفقرة والاختبار باستخدام برمجية Bilog-Mg:

باستخدام طريقة الأرجحية العظمى تم حساب الخطأ المعياري للقياس (Standard Error of Measurement (SEM)) للنماذج اللوجستية.

وتم الحصول على دالة معلومات كل فقرة في المرحلة الثالثة (Ph3) من مراحل استخراج النتائج مع رسم لدالة معلومات كل فقرة بالإضافة إلى رسم لدالة معلومات الاختبار مع الخطأ المعياري للقياس.

تاسعاً: حساب الدقة النسبية والفاعلية النسبية للاختبار (Relative Precision & Relative Efficiency):

تم حساب كل من الدقة النسبية للاختبار والفاعلية النسبية وفق الخطوات الآتية:

1. حساب دالة معلومات الاختبار عند كل مستوى قدرة لكل من النماذج الثلاثة.

2. معادلة تقديرات قدرات الأفراد للاختبار وفق النموذج ثلاثي المعلَمة إلى تقديرات القدرة للاختبارين وفق النموذجين ثنائي المعلَمة وأُحادي المعلَمة باستخدام الرتب المئينية.

3. حساب الدقة النسبية للاختبار من خلال قسمة دالة المعلومات للاختبار وفق النموذج ثلاثي المعلَمة على دالة المعلومات للاختبار وفق النموذج أُحادي المعلَمة عند كل مستوى قدرة، كذلك من خلال قسمة دالة المعلومات للاختبار وفق النموذج ثنائي المعلَمة على دالة المعلومات للاختبار وفق النموذج أُحادي المعلَمة عند كل مستوى قدرة.

4. حساب الفاعلية النسبية للاختبار من خلال قسمة دالة المعلومات للاختبار وفق النماذج اللوجستية الثلاثة والمستل من الاختبار الأصلي ويمثل الفقرات الأكثر صعوبة على دالة معلومات الاختبار الكلي عند مستوى كل قدرة، كذلك من خلال قسمة دالة المعلومات للاختبار وفق النماذج اللوجستية الثلاثة والمستل من الاختبار الأصلي ويمثل الفقرات الأكثر سهولة على دالة معلومات الاختبار الكلي عند مستوى كل قدرة.

عاشراً: حساب المؤشرات الإحصائية الوصفية للاختبار الكلي (89 فقرة) والاختبار بصورته النهائية (82 فقرة):

- تم حساب الرتب المئينية ومقاييس النزعة المركزية ومقاييس التشتت والتفلطح والإلتواء والتوزيع التكراري للعلامات الخام على الاختبار الكلي (89 فقرة)، والاختبار بصورته النهائية (82 فقرة)، والمكوّن من الفقرات التي طابقت النموذج ثلاثي المعلمة لمجموعة الطلبة الجامعيين وطلبة المرحلة الثانوية وطلبة المرحلة الأساسية العليا كلٌ على انفراد:

حادي عشر: التحقق من ميزات (خصائص) النموذج:

تم التحقق من ميزات النموذج من خلال:

1. التحقق من خاصية اللاتغير لمقدرات معالم الفقرة (Invariance of Item Parameter Estimates).

تم اختيار عينة عشوائية من الذكور من عينة الدراسة عددها 488 طالب، وعينة عشوائية من الإناث عددها 488 طالبة، تمت قسمة كل عينة إلى قسمين متساويين بحيث تكونت المجموعات الآتية: ذكور1، ذكور2، إناث1، إناث2 بلغ عدد كل مجموعة 244 فرد.

تم رسم تمثيل بياني لمعالم الفقرة على النماذج اللوجستية الثلاثة للمجموعات الآتية: (ذكور 1، ذكور 2)، (ذكور 1، إناث 1)، (ذكور 1، إناث 2)، (إناث 1، إناث 2). كما تم حساب معامل ارتباط سبيرمان (الرتب) لكل مجموعة من المجموعات السابقة.

كذلك قامت الباحثة بمقارنة الأوساط الحسابية والانحرافات المعيارية لمعلم الصعوبة في كل مجموعة وحساب دلالة الفروق بينهم باستخدام تحليل التباين الأحادي.

2. التحقق من خاصية اللاتغير لمقدرات القدرة (Invariance of Ability Parameter).

تمت قسمة فقرات كل نموذج إلى قسمين متساويين: القسم الأول يمثل الفقرات السهلة والقسم الثاني يمثل الفقرات الصعبة. تم تطبيق واستخراج النتائج لأفراد الدراسة على هذين النوعين من الفقرات، وتم رصد تقديرين لمعلّم القدرة لكل مفحوص على مجموعتي الفقرات السهلة والصعبة. مُثلت هذه التقديرات بيانياً، كما تم الحصول على معامل الارتباط سبيرمان بين التقديرين.

كما قامت الباحثة بمقارنة المتوسطين الحسابيين لقدرة المفحوصين على الاختبارين السهل والصعب باستخدام اختبار ت.

الفصل الرابع

النتائج

يتناول هذا الفصل نتائج الدراسة في ضوء الهدف الذي تسعى إليه وهو بناء اختبار لمهارات التفكير الناقد باستخدام نظرية استجابة الفقرة، وفيما يأتي عرض لهذه النتائج:

التحقق من الافتراضات التي تتطلبها نظرية استجابة الفقرة:

قامت الباحثة بالتحقق من افتراضات نظرية استجابة الفقرة في فقرات الاختبار الكلي (89 فقرة) وهي:

- افتراض أُحادية البعد (Unidimensionality):

تم التحقق من هذا الافتراض بعدة مؤشرات هي:

1. المؤشرات التي اعتمدت على التحليل العاملي للمكونات الأساسية (Principal Component).

تم إجراء تحليل عاملي لبيانات الاختبار الكلي المتعلقة باستجابات 1174 فرداً عن 89 فقرة تمثل الاختبار الكلي، أفرز التحليل 28 عاملاً فسر العامل الأول منها 12.12% من التباين، وفسرت جميع العوامل 53.24% من التباين الكلي، وكانت قيمة الجذر الكامن للعامل الأول 10.79 والثاني 2.76. تم إجراء تحليل عاملي من الرتبة الثانية (Second Order Factor Analysis) متبوعاً بإدارة المحاور المتعامدة (Varimax Rotation)، حيث أفرز التحليل ستة عوامل فسر العامل الأول منها 24.34% من التباين، وفسرت جميع العوامل 44.65% من التباين. يبيّن الجدول 9 قيم الجذر الكامن ونسبة التباين المفسر لكل عامل، كذلك نسبة التباين المفسّر التراكمية، وعدد فقرات الاختبار المتشبعة على كل عامل.

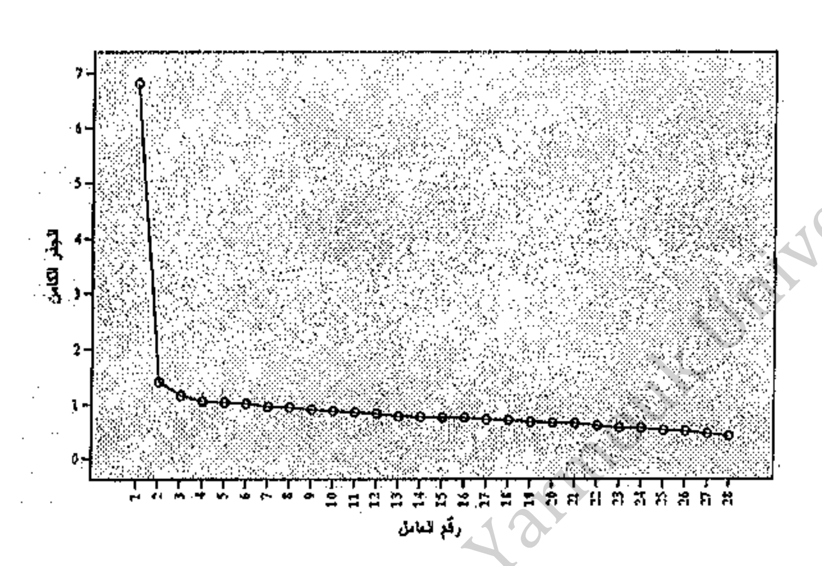
جدول 9

نتائج التحليل العاملي من الرتبة الثانية للاختبار الكلي (89 فقرة)

رقم العامل	الجذر الكامن	نسبة التباين المفسر %	نسبة التباين المفسر التراكمية%	عدد الفقرات المشبعة بالعامل
1	6.81	24.34	24.34	46
2	1.42	5.06	29.39	20
3	1.17	4.16	33.56	11
4	1.06	3.78	37.34	5
5	1.04	3.70	41.04	5
6	1.01	3.62	44.65	2

يتبين من الجدول 9 أن قيم الجذر الكامن للعامل الأول 6.81 ويفسر ما نسبته 24.3% من التباين الكلي، وبلغت قيمة الجذر الكامن الأول إلى الجذر الكامن الثاني 4.8، كما بلغت قيمة الفرق بين الجذرين الكامنين الأول والثاني إلى الفرق بين الجذرين الكامنين الثاني والثالث 21.6. وجميع هذه المؤشرات تدل على تحقق افتراض أُحادية البعد للمقياس بصورته الكلية أي أن المقياس يقيس سمة التفكير الناقد والتي تتكوّن من عدة مهارات للتفكير الناقد.

ويتعزز افتراض أُحادية البعد من خلال اختبار فحص العوامل (Scree Plot) الذي يظهر في الشكل 1. ويتبين من الشكل أن الجذر الكامن للعامل الأول يتميز بشكل واضح عن الجذور الكامنة لبقية العوامل.



الشكل 1: التمثيل البياني لقيم الجذور الكامنة للعوامل المكونة للاختبار على البيانات الكلية

2. المؤشرات التي اعتمدت معامل الثبات:

تم حساب معامل ارتباط بايسيريل النقطي بين الأداء على الفقرة والأداء على الاختبار الكلي بعد حذف الفقرة (Corrected Item Total Correlation). ويبين الجدول 10 قيم تلك المعاملات للاختبار الكلي (89 فقرة). ويلاحظ من الجدول أن جميع معاملات الارتباط تزيد على 0.20. وتدل هذه القيم لمعاملات الارتباط وفق محك ننلي (Nunnally, cited in Hattie, 1985) على أن الفقرات تقيس بعداً واحداً. هذا وقد بلغت قيمة KR-20 لبيانات الاختبار الكلي 0.91.

جدول 10

معاملات الارتباط المصححة للفقرات مع الاختبار الكلي (89 فقرة)

الفقرة	معامل الارتباط المصحح للفقرة بالاختبار	الفقرة	معامل الارتباط المصحح للفقرة بالاختبار
1	0.28	46	0.32
2	0.34	47	0.25
3	0.35	48	0.42
4	0.30	49	0.28
5	0.30	50	0.30
6	0.33	51	0.42
7	0.27	52	0.33
8	0.30	53	0.25
9	0.38	54	0.29
10	0.22	55	0.22
11	0.35	56	0.34
12	0.39	57	0.24
13	0.33	58	0.28
14	0.46	59	0.27
15	0.28	60	0.34
16	0.35	61	0.21
17	0.30	62	0.24
18	0.38	63	0.24
19	0.26	64	0.30
20	0.34	65	0.31
21	0.36	66	0.31
22	0.36	67	0.31
23	0.34	68	0.28
24	0.30	69	0.37
25	0.22	70	0.34
26	0.37	71	0.28
27	0.27	72	0.43
28	0.21	73	0.33
29	0.26	74	0.27
30	0.26	75	0.31
31	0.28	76	0.34
32	0.41	77	0.42
33	0.32	78	0.22
34	0.23	79	0.40
35	0.28	80	0.33
36	0.24	81	0.35
37	0.32	82	0.37
38	0.34	83	0.37
39	0.28	84	0.39
40	0.28	85	0.25
41	0.36	86	0.21
42	0.29	87	0.35
43	0.39	88	0.38
44	0.29	89	0.43
45	0.25		

يتبين من الجدول 10 أن معاملات ارتباط الفقرات مع الاختبار لبيانات الاختبار بصورته الكلية (89 فقرة) تراوحت بين 0.21و0.46.

- افتراض الاستقلال الموضعي (Local Independence):

نظراً لأن افتراض الاستقلال الموضعي يكافئ افتراض أُحادية البعد فإن الباحثة اكتفت بالتحقق من افتراض أُحادية البعد؛ للإشارة على تحقق افتراض الاستقلال الموضعي (Hambleton, 1998; Hambleton & Swaminathan, 1985).

- عامل السرعة (Speededness):

تأكدت الباحثة من أن عامل السرعة لم يلعب دوراً في الإجابة عن فقرات الاختبار، بمعنى أن إخفاق الأفراد في الاختبار يعود إلى انخفاض قدراتهم وليس إلى تأثير عامل السرعة على الإجابة، وذلك عن طريق إعطاء الوقت الكافي للطلبة للإجابة عن فقرات الاختبار، ولم يشتكِ أي طالب في أثناء التطبيق من ضيق الوقت وعدم كفايته. علماً بأن زمن الاختبار تم تقديره في ضوء المعلومات التي توفرت للباحثة في مرحلة تجريب الاختبار على عينة استطلاعية.

النتائج المتعلقة بصدق الاختبار الكلي:

تم الاطمئنان إلى أن الاختبار يقيس مهارات التفكير الناقد التي صمم لقياسها من خلال الخطوات التي اتبعت في بناء الاختبار وبالاستناد إلى آراء المحكمين التي توافقت حول المهارات المقاسة من جهة، وملاءمة الفقرات لقياس تلك المهارات من جهة أخرى. واستناداً إلى الأدب السابق فقد تم تبني ثلاث فرضيات: تنص الأولى على عدم وجود فروق في التفكير الناقد بين الذكور والإناث، وتنص الثانية على أن التفكير الناقد يتحسن بازدياد المستوى التعليمي، وتنص الثالثة على وجود فروق في التفكير الناقد بين ذوي التحصيل المرتفع والمتدني.

واستخدمت الباحثة اختبار ت للمقارنة بين المتوسطين الحسابيين للعلامات على الاختبار لعينة من طلبة جامعة اليرموك عددهم 24 طالباً وطالبة تم اختيارهم عشوائياً من عينة الدراسة، مع طلبة كلية الطب من عينة الدراسة، وعددهم 24 طالباً وطالبة.

كما استخدمت الباحثة اختبار ت للمقارنة بين المتوسطين الحسابيين للعلامات على الاختبار لعينة من طلبة الصف العاشر من ذوي التحصيل العالي ممثلة بطلبة مدارس الملك عبد الله الثاني للتميز عددهم 39 طالباً وطالبة، وبين عينة من طلبة الصف العاشر من ذوي التحصيل المتدني الذين معدلاتهم المدرسية دون 65، يبيّن الجدول 11 نتائج اختبار ت للمقارنة بين المتوسطين الحسابيين للعلامات على الاختبار الكلي لكل من عينة طلبة جامعة اليرموك وطلبة جامعة العلوم والتكنولوجيا (كلية الطب)، وطلبة الصف العاشر من ذوي التحصيل العالي (مدارس التميّز) والتحصيل المتدني.

جدول 11

المتوسطات الحسابية والانحرافات المعيارية وقيمة ت لدلالة الفرق بينهما وفق المستوى التعليمي

المستوى التعليمي	العدد	المتوسط	الانحراف	قيمة ت المحسوبة	درجة الحرية	الدلالة الإحصائية
طلبة جامعة اليرموك	24	43.542	12.07			
طلبة جامعة العلوم والتكنولوجيا (ط.ب)	24	66.417	13.49	6.191	46	0.000
طلبة صف عاشر/تحصيل متدني	39	36.03	11.884			
طلبة صف عاشر/تحصيل عالي	39	46.85	12.171	-3.972	76	0.000

واستخدمت الباحثة تحليل التباين الثنائي (2×3)، وذلك؛ لاستخراج دلالة الفروق على الاختبار بين المستويات التعليمية المختلفة: أساسي عالي ممثلاً بطلبة الصفين التاسع والعاشر، ثانوي ممثلاً بطلبة الصف الأول الثانوي، والمستوى الجامعي، ولاستخراج دلالة الفرق في الأداء، على الاختبار بين الذكور والإناث. وتم اعتماد النوع الثالث (Type III) لاستخراج مجموع المربعات الخاص بالتحليل بسبب عدم اتزان الخلايا وعدم وجود خلايا فارغة، ويتم استخدامه أيضاً من قبل البرنامج بشكل افتراضي، يبيّن الجدول 12 نتائج هذا التحليل.

جدول 12

نتائج تحليل التباين لأثر متغيري المستوى التعليمي والجنس على أداء أفراد عينة الدراسة على الاختبار الكلي (89 فقرة)

مصدر التباين	مجموع المربعات	درجة الحرية	متوسط المربعات	قيمة ف المحسوبة	الدلالة الإحصائية	الدلالة العملية
الجنس	126.151	1	126.151	0.717	0.397	0.001
المستوى التعليمي	32443.064	2	16221.532	92.136	0.000	0.136
الجنس*المستوى التعليمي	4448.416	2	2224.208	12.633	0.000	0.021
الخطأ	205638.066	1168	176.060			
الكلي	241533.380	1173				

أظهرت نتائج تحليل التباين عدم وجود أثر ذي دلالة للجنس في الأداء على الاختبار حيث كان المتوسط الحسابي لعلامات الذكور 44.261 ولعلامات الإناث 44.896. بينما يوجد أثر للمستوى التعليمي حيث كان المتوسط الحسابي لعلامات طلبة المرحلة الأساسية على الاختبار 38.82 وهو الأدنى، في حين كان 46.74 لطلبة المرحلة الثانوية، و55.02 لطلبة المرحلة الجامعية وهو الأعلى.

وعند إجراء المقارنات البعدية للفروق على الأداء بين المجموعات (المستويات التعليمية المختلفة) تم استخدام اختبار جيمس هاول (Games-Howell) لعدم تجانس التباين. يبيّن الجدول 13 هذه المقارنات.

جدول 13

الفروق بين المتوسطات بطريقة جيمس هاول للمقارنة بين المتوسطات الحسابية لأداء الطلبة في المستويات التعليمية المختلفة على الاختبار الكلي (89 فقرة)

المستوى التعليمي		المرحلة الأساسية العليا	المرحلة الثانوية	المرحلة الجامعية
	المتوسط	38.820	46.744	55.017
المرحلة الأساسية العليا	38.820			
المرحلة الثانوية	46.744	*7.924		
المرحلة الجامعية	55.017	*16.197	*8.272	

* احتمالية الخطأ < 0.01.

يتبين من الجدول 13 أن هناك فروقاً ذات دلالة إحصائية (α= 0.01) ولصالح المستوى التعليمي الأعلى.

كما قامت الباحثة بتطبيق اختبار كاليفورنيا للتفكير الناقد على شعبتين من شعب الصف الأول الثانوي تم اختيارهما عشوائياً من شعب الدراسة، تم حساب معامل الارتباط بين علامات الطلبة على اختبار كاليفورنيا والاختبار الكلي (89 فقرة) حيث كان معامل الارتباط 0.723.

كما تم حساب معامل الارتباط بين علامات نفس طلبة الشعبتين على الاختبار ومعدل علاماتهم المدرسية، حيث بلغ معامل الارتباط 0.834.

النتائج المتعلقة بمطابقة الفقرات للنماذج الثلاثة:

قامت الباحثة بإدخال البيانات في ذاكرة الحاسوب وأخضعت البيانات الخاصة باستجابة أفراد عينة الدراسة 1174 فرداً عن فقرات الاختبار 89 فقرة للتحليل بعد استبعاد استجابة 14 فرداً أظهرت استجابات بعضهم عدم الجدية، وآخرون من طلبة الجامعة تغيبوا عن أحد قسمي الاختبار. ويتم استخراج النتائج من برنامجBilog- Mg وفق المراحل الآتية:

- المرحلة الأولى: يتم عرض نتائج خاصة بالفقرة والاختبار وفق النظرية التقليدية في القياس حيث تعرض مخرجات هذه المرحلة معامل الصعوبة للفقرة من خلال عرض النسبة المئوية لعدد الأفراد الذين أجابوا عن الفقرة إجابة صحيحة، وتقدير آخر لصعوبة الفقرة باللوجيت، ومعامل ارتباط الفقرة مع الاختبار بدلالة معامل الارتباط بايسيريل النقطي.
- المرحلة الثانية: تُظهر النتائج في هذه المرحلة تقديرات معالم الفقرة، والأخطاء المعيارية لكل معلَم، وإحصائي حسن المطابقة.

- **المرحلة الثالثة**: تُظهر النتائج في هذه المرحلة علامة الأفراد على الاختبار والقدرة ودالة معلومات الفقرة والقيمة العظمى لهذه الدالة، ورسم دالة معلومات الفقرة والاختبار.

تم استخدام البرنامج لتحليل البيانات الخام ثلاث مرات بشكل مستقل لاختيار الفقرات المناسبة لكل نموذج من النماذج اللوجستية الثلاثة لنظرية استجابة الفقرة، في المرة الأولى لاختيار الفقرات المطابقة لنموذج أُحادي المعلَمة بالاعتماد على استجابات أفراد عينة الدراسة الكلية (1174 فرداً) على الاختبار الكلي (89 فقرة) تبيّن عدم مطابقة استجابات 75 فرداً، تراوحت علاماتهم الخام بين 27 و 87، في حين تراوحت قدراتهم بين −1.247 و4.

وأفرزت نتائج التحليل عند استخدام البرنامج لأول مرة على البيانات الخام وفق نموذج ثنائي المعلَمة عدم مطابقة استجابات 74 فرداً تراوحت علاماتهم الخام بين 27 و84، في حين تراوحت قدراتهم بين −1.326 و 4.

وفيما يتعلق بنموذج ثلاثي المعلَمة أفرزت نتائج تحليل الاستجابات عدم مطابقة استجابة 98 فرداً تراوحت علاماتهم الخام بين 7 و 64، في حين تراوحت قدراتهم بين −4 و1.226.

تم حذف استجابة الأفراد الذين لم تتطابق استجاباتهم مع توقعات النموذج، وأعيد التحليل لاختبار مدى مطابقة الفقرات لكل نموذج، حيث أظهرت نتائج التحليل في المرحلة الثانية والخاصة بمطابقة الفقرات ومن خلال اختبار كاي تربيع عند مستوى الدلالة ($\alpha = 0.01$) عدم مطابقة 39 فقرة لنموذج أُحادي المعلَمة، وعدم مطابقة 19 فقرة لنموذج ثنائي المعلمة، في حين كانت هناك سبع فقرات فقط لم تطابق النموذج ثلاثي المعلمة.

بعد ذلك تم حذف الفقرات غير المطابقة لكل نموذج، وأعيد التحليل للحصول على تقديرات نهائية لكل من معالم الفقرات وقدرات الأفراد على النماذج الثلاثة.وبذا تم الحصول على الاختبارات الآتية: الاختبار وفق النموذج أُحادي المعلمة ويتكون من 50 فقرة، والاختبار وفق النموذج ثنائي المعلمة ويتكون من 70 فقرة، والاختبار وفق النموذج ثلاثي المعلمة ويتكون من 82 فقرة. حيث شكلت فقرات الاختبار وفق النموذج أُحادي المعلمة 56% من فقرات الاختبار الكلـــي (89 فقرة)، وشكلت فقرات الاختبار وفق النموذج ثنائي المعلمة 79% من فقرات الاختبار الكلي، والنسبة الأكبر كانت للاختبار وفق النموذج ثلاثي المعلمة إذ شكل نسبة 92% من فقرات الاختبار الكلي.

تم تقدير معالم الفقرة بطريقة الأرجحية العظمى الهامشية (MML) من خلال استخدام برمجية Biolg-Mg والتي تعطي تقديرات دقيقة عن طريق إعادة متعاقبة لعمليات التقدير. حيث أظهرت نتائج تقدير معالم الفقرة للاختبار أن معلم الصعوبة للفقرات يتراوح بين −0.924 و 1.602 بمتوسط حسابي 0.15، عند استخدام النموذج أُحادي المعلمة وبين − 1.607 و 2.131 بمتوسط حسابي −0.03 عند استخدام النموذج ثنائي المعلمة، وبين −1.242 و 1.642 بمتوسط حسابي 0.59 عند استخدام النموذج ثلاثي المعلمة، وتراوحت قيم تقديرات معلَم التمييز لفقرات الاختبار بين 0.245 و0.761 بمتوسط حسابي 0.48 في النموذج ثنائي المعلمة، وبين 0.37 و2.05 بمتوسط حسابي 1.010 في النموذج ثلاثي المعلمة، وبينت النتائج أن قيم معلم التخمين للاختبار وفق النموذج ثلاثي المعلمة تراوحت بين 0.10 و 0.46 بمتوسط حسابي 0.25.

وبدراسة الأرقام التسلسلية لفقرات الاختبارات وفق النماذج اللوجستية الثلاثة تبين أن هناك 49 فقرة مشتركة، امتازت كل فقرة منها بمطابقة توقعات كل نموذج من النماذج اللوجستية الثلاثة لنظرية استجابة الفقرة، تراوحت تقديرات معالم الصعوبة للفقرات المشتركة وفق النموذج أُحادي المعلمة بين –0.924 و 1.602، وبين –0.983 و 2.131 وفق النموذج ثنائي المعلمة، وبين –0.108 و1.63 وفق النموذج ثلاثي المعلَمة. تراوحت قيم معلم التمييز لهذه الفقرات بين 0.326 و0.649 وفق النموذج ثنائي المعلمة وبين 0.435 و 1.597 وفق النموذج ثلاثي المعلمة، وتراوحت قيم معلم التخمين لفقرات الاختبار وفق النموذج ثلاثي المعلمة بين 0.148 و 0.357.

امتازت هذه الفقرات المشتركة (49 فقرة) بأنها غطت المجال الكلي للمهارات الأساسية التي تم بناء الاختبار على أساسها، وهي: التحليل وضمت خمس فقرات، والتفسير وضمت تسع فقرات، والتقويم وضمت أربع فقرات، والشرح وضمت أربع فقرات، والاستدلال وضمت 16 فقرة، والتبرير الإحصائي والاحتمالي وضمت ثلاث فقرات، والتعرف على الافتراضات المنصوص وغير المنصوص عليها وضمت خمس فقرات، والنظر إلى الأشياء بعمق وتوسع (التركيز) وضمت ثلاث فقرات.

وقد قامت الباحثة بعد ذلك باستبعاد استجابة أي فرد تم استبعاد استجابته على الاختبار وفق أحد النماذج على الأقل، حيث بلغ عدد الأفراد المستبعدين 110 أفراد، وأعادت التحليل مرة أخرى ثلاث مرات بشكل مستقل لحساب التقديرات النهائية لكل من معالم الفقرات وقدرات الأفراد على النماذج اللوجستية الثلاثة للبيانات الخاصة باستجابات نفس الأفراد وعددهم 1064 فرداً عن الفقرات المشتركة وعددها 49 فقرة، وبذا تكون الصورة المناسبة للاختبار لمقارنة النماذج اللوجستية الثلاثة من أجل تحديد النموذج الأنسب مكونة من 49 فقرة طابقت توقعات النماذج اللوجستية الثلاثة (الأحادي، والثنائي، والثلاثي).

النتائج المتعلقة بمؤشرات الثبات:

استخدمت الباحثة برمجيةBilog -Mg؛ لاستخراج الثبات النظري والتجريبي للفقرات المشتركة (49 فقرة) للاختبار وفق النماذج اللوجستية الثلاثة لنظرية استجابة الفقرة، كذلك استخدمت برمجية SPSS لحساب معامل الاتساق الداخلي باستخدام معادلة KR-20 ويبيّن الجدول 14 هذه النتائج.

جدول 14

معاملات الثبات النظري والتجريبي للفقرات المشتركة والاتساق الداخلي للاختبار وفق النماذج اللوجستية الثلاثة لنظرية استجابة الفقرة

النموذج	معامل الثبات		
	النظري	التجريبي	الاتساق الداخلي
أُحادي المعلَمة	0.856	0.861	0.857
ثنائي المعلَمة	0.860	0.867	0.857
ثلاثي المعلَمة	0.878	0.908	0.857

يتبين من الجدول 14 أن أعلى معاملي ثبات (النظري والتجريبي) للاختبار يتحقق عند استخدام النموذج ثلاثي المعلَمة، ويتميز معامل الثبات التجريبي والذي يعتمد على كيفية تقدير علامات القدرة بأنه أعلى من معامل الثبات النظري والذي يعتمد على تقديرات معالم الفقرة للاختبار وفق النماذج الثلاثة، كما لوحظ أن هناك شبه تطابق بين قيمتي معاملي الثبات النظري للاختبار وفق النموذجين أُحادي وثنائي المعلمة من طرف ومعامل الاتساق الداخلي للاختبار الممثل بالفقرات المشتركة من طرف آخر.

النتائج المتعلقة بتقديرات معالم الفقرات المشتركة للاختبار وفق النماذج اللوجستية الثلاثة:

تم تقدير معالَم الفقرة بطريقة الأرجحية العظمى الهامشية (MML) من خلال استخدام برمجية Bilog- Mg والتي تعطي تقديرات دقيقة عن طريق إعادة متعاقبة لعمليات التقدير.

يبيّن الجدول 15 قيم معلَم صعوبة الفقرات والخطأ المعياري في تقدير هذا المعلَم للفقرات المشتركة للاختبار وفق النموذج اللوجستي أُحادي المعلَمة.

جدول 15

قيم معلَم صعوبة الفقرات والخطأ المعياري في تقدير معلَم الصعوبة للفقرات المشتركة للاختبار وفق النموذج اللوجستي أُحادي المعلَمة

رقم الفقرة في الاختبار	معلم الصعوبة	الخطأ المعياري لمعلم الصعوبة	رقم الفقرةفي الاختبار	معلم الصعوبة	الخطأ المعياري لمعلم الصعوبة
1	-0.897	0.088	45	-0.545	0.085
2	0.868	0.092	46	-0.209	0.085
3	0.103	0.088	47	-0.708	0.086
4	-0.932	0.089	49	-0.187	0.084
8	-0.395	0.085	51	0.213	0.089
11	0.403	0.088	55	1.658	0.098
13	1.003	0.092	56	0.448	0.088
16	1.194	0.095	57	-0.833	0.086
17	-0.285	0.085	58	-0.039	0.084
18	0.471	0.089	60	0.308	0.087
20	1.353	0.098	62	0.911	0.089
21	-0.258	0.088	63	0.563	0.085
22	0.874	0.091	64	-0.417	0.086
23	-0.495	0.087	66	-0.384	0.086
24	-0.874	0.088	67	-0.788	0.088
27	0.131	0.085	69	0.814	0.092
29	0.551	0.087	71	1.313	0.095
33	-0.456	0.086	74	0.125	0.084
35	0.725	0.088	75	-0.985	0.089
39	0.437	0.086	76	-0.484	0.087
40	1.136	0.092	80	-0.127	0.087
41	0.026	0.087	81	1.015	0.093
42	0.109	0.086	82	-0.545	0.088
43	-0.473	0.089	85	0.454	0.085
44	0.241	0.086			

يلاحظ من الجدول 15 أن قيم معلَم الصعوبة للفقرات يتراوح بين -0.985 و 1.658بمتوسط حسابي 0.125. في حين تراوحت قيم الخطأ المعياري لتقدير معلم الصعوبة بين 0.084 و 0.098.

ويبيّن الجدول 16 معالم الفقرات والخطأ المعياري لتقدير هذه المعالم للفقرات المشتركة للاختبار وفق النموذج اللوجستي ثنائي المعلَمة.

جدول 16

قيم معالم الفقرة والخطأ المعياري لتقدير المعالم للفقرات المشتركة للاختبار وفق النموذج اللوجستي ثنائي المعلَمة

رقم الفقرة في الاختبار	معلم الصعوبة	الخطأ المعياري لمعلم الصعوبة	معلم التمييز	الخطأ المعياري لمعلم التمييز	رقم الفقرة في الاختبار	معلم الصعوبة	الخطأ المعياري لمعلم الصعوبة	معلم التمييز	الخطأ المعياري لمعلم التمييز
1	-0.933	0.122	0.424	0.047	45	-0.652	0.125	0.348	0.048
2	0.739	0.100	0.532	0.048	46	-0.220	0.093	0.408	0.046
3	0.068	0.074	0.540	0.049	47	-0.936	0.164	0.310	0.042
4	-0.861	0.103	0.500	0.056	49	-0.205	0.101	0.371	0.045
8	-0.414	0.095	0.419	0.048	51	0.128	0.067	0.631	0.056
11	0.342	0.083	0.517	0.048	55	2.050	0.254	0.346	0.041
13	0.950	0.122	0.472	0.045	56	0.404	0.090	0.490	0.047
16	1.095	0.127	0.492	0.047	57	-1.104	0.177	0.312	0.041
17	-0.306	0.097	0.399	0.046	58	-0.032	0.107	0.350	0.041
18	0.370	0.076	0.562	0.046	60	0.272	0.085	0.492	0.048
20	1.115	0.111	0.563	0.049	62	1.121	0.170	0.351	0.042
21	-0.238	0.068	0.585	0.052	63	0.819	0.165	0.297	0.039
22	0.784	0.104	0.501	0.046	64	-0.425	0.092	0.438	0.050
23	-0.471	0.084	0.488	0.051	66	-0.407	0.098	0.411	0.045
24	-0.880	0.116	0.442	0.050	67	-0.797	0.110	0.439	0.047
27	0.143	0.096	0.411	0.046	69	0.636	0.091	0.586	0.053
29	0.612	0.119	0.396	0.044	71	1.339	0.160	0.434	0.045
33	-0.470	0.096	0.429	0.044	74	0.175	0.112	0.341	0.041
35	0.810	0.132	0.392	0.044	75	-0.958	0.116	0.464	0.051
39	0.494	0.111	0.390	0.042	76	-0.472	0.087	0.469	0.051
40	1.298	0.166	0.381	0.042	80	-0.130	0.077	0.505	0.051
41	0.009	0.077	0.503	0.046	81	0.923	0.114	0.498	0.047
42	0.115	0.092	0.421	0.044	82	-0.471	0.072	0.580	0.057
43	-0.389	0.062	0.672	0.062	85	0.617	0.141	0.322	0.041
44	0.269	0.101	0.400	0.044					

يلاحظ من الجدول 16 أن قيم معالم الصعوبة تتراوح بين -1.104 و 2.05 بمتوسط حسابي 0.121. وتراوحت قيم الخطأ المعياري لتقدير معلم الصعوبة بين 0.062 و0.254 في حين تراوحت قيم معالم التمييز بين 0.297 و 0.672 بمتوسط حسابي 0.449. وتراوحت قيم الخطأ المعياري لتقدير معلم التمييز بين 0.039 و0.062.

يبيّن الجدول 17 قيم معالم الفقرات والخطأ المعياري لتقدير هذه المعالم للفقرات المشتركة للاختبار وفق النموذج اللوجستي ثلاثي المعلمة.

جدول 17

قيم معالم الفقرة والخطأ المعياري لتقدير المعالم للفقرات المشتركة للاختبار وفق النموذج اللوجستي ثلاثي المعلَمة

رقم الفقرة في الاختبار	معلم الصعوبة	الخطأ المعياري لمعلم الصعوبة	معلم التمييز	الخطأ المعياري لمعلم التمييز	معلم التخمين	الخطأ المعياري لمعلم التخمين
1	-0.162	0.300	0.526	0.077	0.239	0.084
2	1.123	0.081	1.262	0.204	0.195	0.024
3	0.671	0.132	0.872	0.141	0.217	0.045
4	0.180	0.251	0.735	0.139	0.347	0.074
8	0.778	0.171	0.834	0.178	0.359	0.049
11	0.902	0.118	0.903	0.150	0.209	0.039
13	1.284	0.083	1.335	0.236	0.207	0.022
16	1.324	0.082	1.369	0.239	0.184	0.021
17	0.793	0.197	0.727	0.157	0.323	0.056
18	0.635	0.113	0.742	0.089	0.101	0.037
20	1.257	0.085	1.067	0.163	0.122	0.023
21	0.433	0.148	0.884	0.134	0.243	0.051
22	1.164	0.097	1.018	0.167	0.183	0.029
23	0.469	0.191	0.795	0.145	0.309	0.059
24	-0.126	0.291	0.550	0.086	0.241	0.084
27	1.118	0.103	1.230	0.242	0.337	0.028
29	1.282	0.102	1.177	0.227	0.279	0.027
33	0.150	0.236	0.558	0.080	0.191	0.070
35	1.361	0.117	0.993	0.215	0.244	0.030
39	1.121	0.155	0.714	0.143	0.220	0.046
40	1.500	0.124	0.874	0.175	0.178	0.029
41	0.448	0.163	0.656	0.090	0.148	0.053
42	0.904	0.161	0.751	0.143	0.256	0.048
43	0.272	0.147	0.973	0.158	0.255	0.054
44	1.111	0.124	0.947	0.192	0.292	0.036
45	1.074	0.120	1.289	0.294	0.473	0.028
46	0.610	0.226	0.623	0.119	0.249	0.066
47	0.081	0.370	0.423	0.073	0.244	0.085
49	1.005	0.169	0.842	0.195	0.355	0.044
51	0.674	0.102	1.053	0.176	0.207	0.038
55	1.798	0.123	1.152	0.255	0.168	0.019
56	0.973	0.119	0.891	0.153	0.214	0.036
57	-0.023	0.398	0.412	0.069	0.259	0.089
58	0.707	0.252	0.518	0.094	0.208	0.067
60	0.912	0.118	0.920	0.168	0.234	0.039
62	1.508	0.104	1.293	0.269	0.255	0.023
63	1.556	0.147	0.896	0.214	0.291	0.031
64	0.613	0.205	0.738	0.157	0.318	0.061
66	0.319	0.257	0.552	0.092	0.218	0.073
67	-0.079	0.275	0.553	0.081	0.227	0.080
69	1.046	0.075	1.385	0.236	0.196	0.024
71	1.466	0.066	1.489	0.300	0.196	0.019
74	0.991	0.230	0.557	0.116	0.234	0.062
75	-0.136	0.300	0.581	0.093	0.272	0.087
76	0.612	0.168	0.673	0.176	0.347	0.051
80	0.526	0.179	0.726	0.129	0.220	0.059
81	1.232	0.096	1.027	0.178	0.171	0.027
82	0.251	0.168	0.612	0.139	0.262	0.064
85	1.444	0.149	0.847	0.201	0.291	0.035

يتبين من الجدول 17 أن قيم معالم الصعوبة تتراوح بين −0.162 و1.798 بمتوسط حسابي 0.799، وتراوحت قيم الخطأ المعياري لتقدير معلم الصعوبة بين 0.075 و0.398، وتراوحت قيم معلم التمييز بين 0.412 و 1.489 بمتوسط حسابي 0.876، وتراوحت قيم الخطأ المعياري لتقدير معلم التمييز بين 0.069 و0.300، وتراوحت قيم معلم التخمين بين 0.101 و 0.473 بمتوسط حسابي 0.245 وتراوحت قيم الخطأ المعياري لتقدير معلم التخمين بين 0.019 و 0.089.

يبيّن الجدول 18 المؤشرات الإحصائية الوصفية لمعالم الفقرات المشتركة للاختبار وفق النماذج اللوجستية الثلاثة.

جدول 18

المؤشرات الإحصائية الوصفية لمعالم الفقرات المشتركة للاختبار وفق النماذج اللوجستية الثلاثة

	النموذج											
	أحادي المعلمة		ثنائي المعلمة				ثلاثي المعلمة					
الإحصائي	معلم الصعوبة	الخطأ المعياري لمعلم الصعوبة	معلم الصعوبة	الخطأ المعياري لمعلم الصعوبة	معلم التمييز	الخطأ المعياري لمعلم التمييز	معلم الصعوبة	الخطأ المعياري لمعلم الصعوبة	معلم التمييز	الخطأ المعياري لمعلم التمييز	معلم التخمين	الخطأ المعياري لمعلم التخمين
أدنى علامة	-0.985	0.084	-1.104	0.062	0.297	0.039	-0.162	0.075	0.412	0.069	0.101	0.019
المتوسط	0.125	0.088	0.121	0.110	0.449	0.047	0.799	0.168	0.876	0.183	0.245	0.047
الوسيط	0.109	0.088	0.115	0.101	0.438	0.046	0.902	0.148	0.872	0.157	0.239	0.045
المنوال	-0.985	0.089	-1.104	0.062	0.297	0.045	-0.162	0.075	0.412	0.069	0.101	0.019
أعلى علامة	1.658	0.098	2.050	0.254	0.672	0.062	1.798	0.398	1.489	0.300	0.473	0.089
الانحراف	0.694	0.003	0.725	0.035	0.087	0.005	0.512	0.078	0.273	0.061	0.068	0.021
الالتواء	0.259	1.318	0.343	1.734	0.379	0.910	-0.248	1.129	0.394	0.423	0.776	0.460
التفلطح	-0.665	1.474	-0.400	4.563	-0.271	1.196	-0.816	0.838	-0.554	-0.392	1.652	-0.931

يتبيّن من الجدول 18 أن المدى الأكبر لمعلم الصعوبة يعود للنموذج اللوجستي ثنائي المعلمة حيث كانت أدنى علامة لمعلم الصعوبة −1.104 وأعلى علامة لهذا المعلم 2.050، وتميّز هذا النموذج بأن مدى التمييز له أصغر من مدى التمييز للنموذج اللوجستي ثلاثي المعلمة، حيث كانت أدنى علامة لمعلم التمييز 0.297 في حين كانت أعلى علامة لهذا المعلم 0.672 وفق النموذج اللوجستي ثنائي المعلمة.

النتائج المتعلقة بتقدير معلَم القدرة:

تم تقدير معلَم القدرة أيضاً باستخدام طريقة الأرجحية العظمى الهامشية من خلال برمجية Bilog-Mg. يبيّن الجدول 19 قدرات الأفراد والأخطاء المعيارية في تقديرها للفقرات المشتركة للاختبار وفق النماذج اللوجستية الثلاثة (الأحادي، والثنائي، والثلاثي) المعلَمة. علماً بأن الباحثة حذفت أدنى ثلاث علامات خام من البيانات (5، 7، 8)؛ بسبب عدم احتساب قيمة الخطأ المعياري لتقدير القدرة من قبل البرنامج Bilog-Mg لهذه العلامات وفق النموذج اللوجستي ثلاثي المعلمة.

جدول 19

العلامة الخام وقدرة الأفراد والأخطاء المعيارية في تقديرها وتكرارات الأفراد ودالة معلومات الاختبار للفقرات المشتركة وفق النماذج اللوجستية الثلاثة (الأُحادي، والثنائي، والثلاثي) المعلّمة

العلامة الخام	التكرار	النموذج: أحادي المعلمة: القدرة	الخطأ المعياري للقدرة	دالة المعلومات	ثنائي المعلمة: القدرة	الخطأ المعياري للقدرة	دالة المعلومات	ثلاثي المعلمة: القدرة	الخطأ المعياري للقدرة	دالة المعلومات
9	9	-1.96	0.50	4.061	-1.98	0.50	3.950	-3.88	5.68	0.031
10	15	-1.78	0.48	4.385	-1.77	0.48	4.351	-3.81	5.82	0.030
11	15	-1.61	0.46	4.686	-1.64	0.46	4.627	-3.54	4.73	0.045
12	25	-1.45	0.45	4.964	-1.43	0.45	5.040	-2.21	1.88	0.284
13	46	-1.30	0.44	5.219	-1.30	0.43	5.296	-2.01	1.82	0.300
14	41	-1.16	0.43	5.451	-1.15	0.42	5.590	-1.53	1.33	0.566
15	59	-1.03	0.42	5.661	-1.02	0.41	5.824	-1.23	1.02	0.963
16	38	-0.89	0.41	5.849	-0.88	0.41	6.057	-1.28	1.05	0.900
17	58	-0.77	0.41	6.015	-0.76	0.40	6.259	-0.76	0.71	2.011
18	67	-0.64	0.40	6.159	-0.67	0.40	6.388	-0.62	0.62	2.583
19	51	-0.52	0.40	6.282	-0.52	0.39	6.575	-0.35	0.52	3.742
20	54	-0.40	0.40	6.384	-0.40	0.39	6.699	-0.21	0.47	4.466
21	45	-0.28	0.39	6.465	-0.28	0.38	6.802	-0.05	0.43	5.522
22	46	-0.17	0.39	6.525	-0.18	0.38	6.873	0.07	0.40	6.402
23	49	-0.05	0.39	6.565	-0.04	0.38	6.928	0.23	0.36	7.720
24	42	0.06	0.39	6.583	0.04	0.38	6.943	0.28	0.35	8.197
25	42	0.18	0.39	6.581	0.15	0.38	6.943	0.40	0.33	9.342
26	33	0.29	0.39	6.559	0.27	0.38	6.917	0.50	0.31	10.325
27	46	0.41	0.39	6.515	0.39	0.38	6.864	0.60	0.30	11.418
28	27	0.53	0.39	6.452	0.54	0.38	6.757	0.70	0.28	12.538
29	32	0.64	0.40	6.367	0.64	0.39	6.669	0.78	0.27	13.437
30	24	0.76	0.40	6.262	0.76	0.39	6.527	0.88	0.26	14.476
31	19	0.89	0.40	6.136	0.90	0.40	6.353	0.96	0.26	15.214
32	26	1.01	0.41	5.989	0.98	0.40	6.228	1.02	0.25	15.676
33	21	1.14	0.41	5.821	1.15	0.41	5.968	1.11	0.25	16.357
34	18	1.27	0.42	5.632	1.27	0.42	5.761	1.19	0.24	16.707
35	14	1.41	0.43	5.422	1.37	0.42	5.567	1.30	0.24	16.989
36	21	1.55	0.44	5.189	1.58	0.44	5.180	1.36	0.24	16.994
37	12	1.70	0.45	4.935	1.70	0.45	4.928	1.44	0.24	16.879
38	13	1.86	0.46	4.658	1.89	0.47	4.563	1.57	0.25	16.228
39	16	2.03	0.48	4.359	2.05	0.49	4.248	1.60	0.25	15.989
40	9	2.21	0.50	4.038	2.23	0.51	3.891	1.75	0.26	14.808
41	11	2.40	0.52	3.891	2.35	0.52	3.674	1.85	0.27	13.851
42	9	2.62	0.55	3.320	2.63	0.56	3.161	1.98	0.29	12.303
43	10	2.88	0.58	2.925	2.88	0.60	2.750	2.14	0.31	10.416
44	12	3.14	0.63	2.505	3.14	0.65	2.370	2.23	0.33	9.217
45	6	3.47	0.70	2.059	3.52	0.73	1.881	2.44	0.37	7.180
46	2	3.89	0.79	1.587	3.95	0.83	1.442	2.51	0.39	6.589

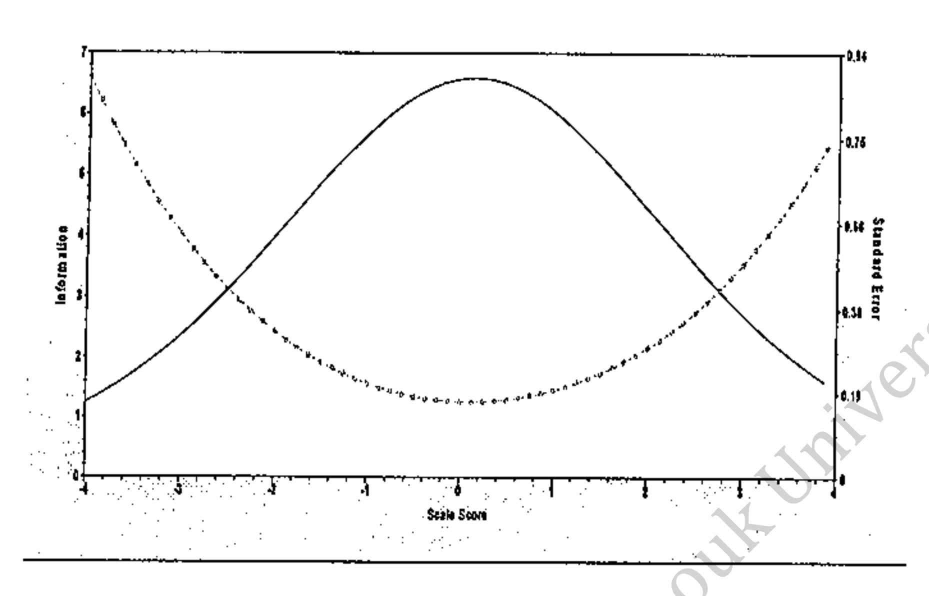
يلاحظ من الجدول 19 أن معلَم القدرة يتراوح بين -1.96 و 3.89 للعلامتين الخام 9 و 46 على الترتيب، وتركزت التكرارات تقريباً في وسط التوزيع للاختبار بصورته النهائية وفق النموذج أُحادي المعلمة.

كما بيّن الجدول أن قيم القدرة للاختبار بصورته النهائية وفق النموذج ثنائي المعلَمة تراوحت بين -1.98 و 3.95 للعلامتين الخام 9 و 46 على الترتيب، أيضاً تركزت التكرارات الأكبر للقدرة تقريباً في وسط التوزيع.

ويُظهر الجدول 19 أيضاً أن قيم القدرة للفقرات المشتركة للاختبار وفق النموذج اللوجستي ثلاثي المعلَمة تراوحت بين -3.88 و 2.51 للعلامتين الخام 9 و 46 على الترتيب كما يلاحظ من الجدول تناقص قيم الخطأ المعياري كلما انتقلنا من مستويات القدرة الدنيا إلى العليا.

ويتبيّن من الجدول أن أول كمية معلومات للاختبار وفق النموذج أُحادي المعلمة كانت 4.061 عند مستوى القدرة -1.96 والمقابلة للعلامة الخام تسعة. وبدأت قيمة كمية المعلومات بالزيادة التدريجية حتى وصلت إلى أعلى كمية معلومات 6.583 عند مستوى القدرة 0.06. بعد ذلك بدأت هذه القيمة بالانخفاض التدريجي إلى أن وصلت أدنى كمية 1.587 عند مستوى القدرة 3.89.

يبيّن الشكل 2 العلاقة بين قيم دالة معلومات الاختبار والخطأ المعياري للتقدير للفقرات المشتركة للاختبار وفق النموذج اللوجستي أُحادي المعلَمة عند كل مستوى من مستويات القدرة.



المنحنى المتصل: يمثل دالة المعلومات

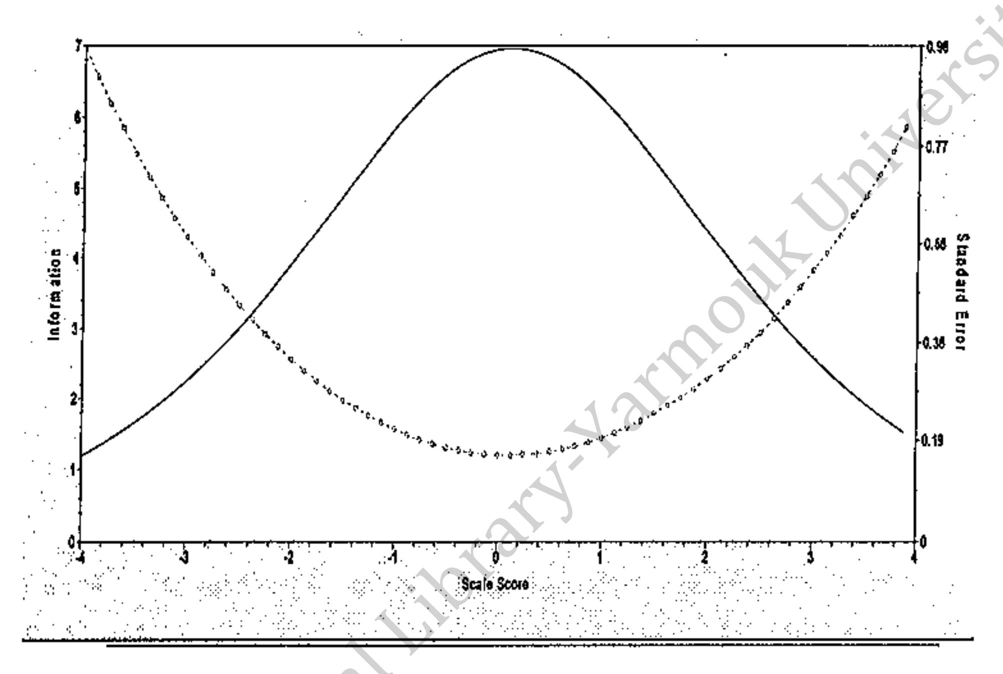
المنحنى المتقطع: يمثل الخطأ المعياري

الشكل 2: دالة معلومات الاختبار والخطأ المعياري للتقدير للفقرات المشتركة للاختبار وفق النموذج اللوجستي أُحادي المعلَمة

ويلاحظ من الشكل أن الاختبار يقدم معلومات أكبر عند القدرات المتوسطة. كما أظهرت نتائج التحليل أن أعلى كمية دالة معلومات يقدمها الاختبار وفق النموذج أُحادي المعلمة هي 6.58 وكانت القيمة العظمى لدالة المعلومات لكل فقرة 0.1432.

ويتبيّن من الجدول أيضاً أن أول قيمة لكمية المعلومات للاختبار وفق النموذج ثنائي المعلمة كانت 3.950 عند مستوى القدرة ‎-1.98 المقابلة للعلامة الخام تسعة، وقدّم النموذج أعلى كمية معلومات عند مستوى القدرتين 0.04 و 0.15، حيث كانت كمية المعلومات 6.943، ثم بدأت قيمة دالة معلومات الاختبار بالانخفاض إلى أن وصلت 1.442 عند مستوى القدرة الأعلى 3.95.

ويبيّن الشكل 3 العلاقة بين قيم دالة معلومات الاختبار والخطأ المعياري للتقدير للفقرات المشتركة للاختبار وفق النموذج اللوجستي ثنائي المعلَمة عند كل مستوى من مستويات القدرة.

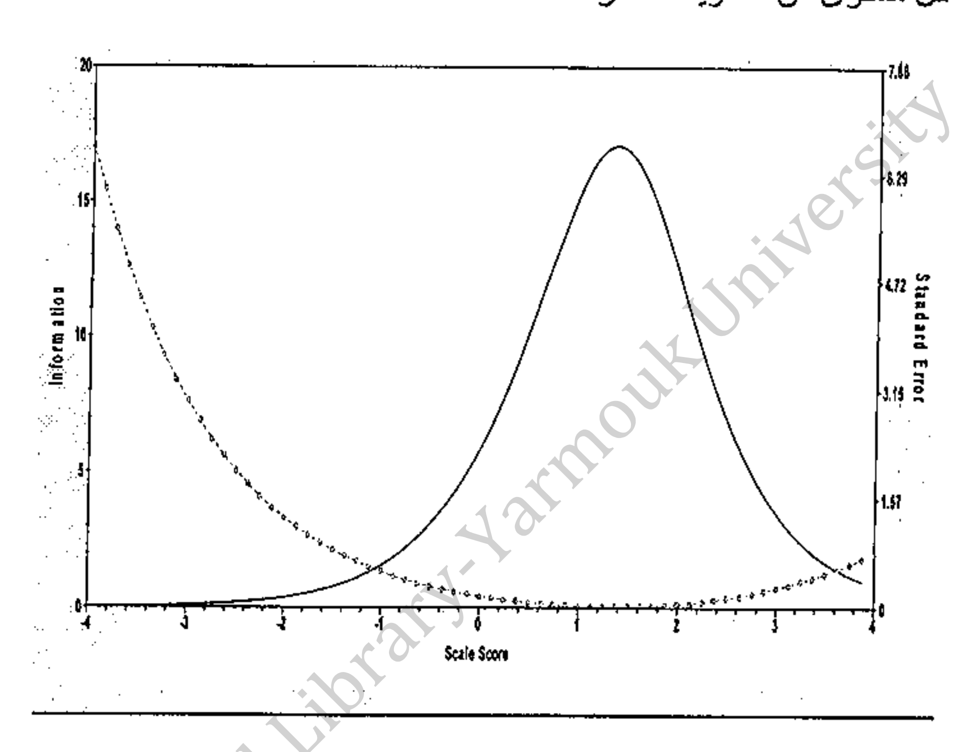


الشكل 3: دالة معلومات الاختبار والخطأ المعياري للتقدير للفقرات المشتركة للاختبار وفق النموذج اللوجستي ثنائي المعلَمة

يلاحظ من الجدول والشكل أن الاختبار يميل إلى إعطاء معلومات أكثر عند الأفراد ذوي القدرات المتوسطة.

كانت أول قيمة لدالة المعلومات قدمها الاختبار وفق النموذج ثلاثي المعلَمة 0.031 عند القدرة –3.88، بدأت بعد ذلك هذه الكمية في الارتفاع التدريجي إلى أن وصلت أعلى قيمة 16.994 المقابلة للقدرة 1.36، ثم بدأت القيم بالانخفاض إلى أن وصلت 6.569 عند مستوى القدرة 2.51.

يبيّن الشكل 4 العلاقة بين قيم دالة معلومات الاختبار وفق النموذج ثلاثي المعلَمة عند كل مستوى من مستويات القدرة.



المنحنى المتصل: يمثل دالة المعلومات

المنحنى المتقطع: يمثل الخطأ المعياري

الشكل 4: دالة معلومات الاختبار والخطأ المعياري للتقدير للفقرات المشتركة للاختبار بصورته النهائية وفق النموذج اللوجستي ثلاثي المعلَمة

كما أظهرت النتائج أن الفقرات قدمت قيماً عظمى لدالة المعلومات للفقرة تراوحت بين 0.745 للفقرة 57 و0.997 للفقرة 71.

ولدى مقارنة كمية المعلومات للاختبار وفق النماذج اللوجستية الثلاثة بافتراض تقسيم متصل السمة إلى ثلاثة أقسام: القسم الأول ويمثل القدرات الدنيا تقريباً لغاية م30، القسم الثاني ويمثل القدرات المتوسطة الواقعة بين م30 و م70، والقسم الثالث يمثل القدرات العليا من م70 إلى م100، يلاحظ من الجدول أن كمية المعلومات التي قدمها الاختبار وفق النموذج أحادي المعلَمة تراوحت بين 4.061 و 6.159 في الثلث الأول لمتصل السمة، في حين تراوحت القيم

بين 3.95 و 6.388 للاختبار وفق النموذج ثنائي المعلَمة، قابلها في الاختبار وفق النموذج ثلاثي المعلَمة قيماً تراوحت بين 0.031 و2.583.

وتراوحت قيم دالة المعلومات بين 6.282 و 6.583 في القسم الثاني لمتصل السمة للاختبار وفق النموذج أُحادي المعلَمة، قابلتها كمية معلومات تراوحت بين 6.575 و 6.943 للاختبار وفق النموذج ثنائي المعلَمة، في حين أظهر الاختبار وفق النموذج ثلاثي المعلَمة أن أدنى كمية معلومات قدمها الاختبار 3.742 وأعلى كمية معلومات قدمها الاختبار 11.418 في القسم الثاني لمتصل السمة.

وبالنظر إلى القسم الثالث والذي يمثل المنطقة الواقعة في الثلث الأخير لمتصل السمة، تراوحت كمية المعلومات للاختبار وفق النموذج أُحادي المعلَمة بين 1.587 و 6.452، في حين تراوحت قيم كمية المعلومات بين 1.442 و 6.757 في الاختبار وفق النموذج ثنائي المعلَمة، وأبدى كل من الاختبارين وفق النموذجين أُحادي وثنائي المعلَمة تناقصاً ملحوظاً في كمية المعلومات في هذا القسم، في حين أظهر الاختبار وفق النموذج ثلاثي المعلَمة خلاف ذلك حيث استمرت كمية المعلومات بالتزايد إذ تراوحت هذه القيم بين 6.569 و16.994.

كما قامت الباحثة بحساب المؤشرات الإحصائية الوصفية لمعلَم القدرة على النماذج اللوجستية الثلاثة، يبيّن الجدول 20 هذه القيم.

جدول 20

المؤشرات الإحصائية الوصفية لمعلَم القدرة والعلامة الخام التي تناظرها للفقرات المشتركة للاختبار وفق النماذج اللوجستية الثلاثة

الإحصائي	العلامة الخام	معلم القدرة		
		أحادي المعلمة	ثنائي المعلمة	ثلاثي المعلمة
أدنى علامة	5	-2.884	-2.976	-4.000
أعلى علامة	46	3.890	3.948	2.507
المتوسط	25.976	0.364	0.359	-0.063
الانحراف	12.022	1.696	1.716	1.973
الإلتواء	-0.013	0.153	0.142	-0.872
التفلطح	-1.181	-0.688	-0.634	-0.298

كما قامت الباحثة بحساب معامل ارتباط بيرسون بين معلَم القدرة والعلامة الخام للفقرات المشتركة للاختبار وفق النماذج اللوجستية الثلاثة، يبيّن الجدول 21 هذه القيم.

جدول 21

معاملات ارتباط بيرسون بين العلامة الخام والقدرة وفق النماذج اللوجستية الثلاثة

النموذج	معامل الارتباط
أحادي المعلمة	0.995
ثنائي المعلمة	0.992
ثلاثي المعلمة	0.906

يتبيّن من الجدول 21 أن معامل الارتباط الأعلى كان للنموذج أُحادي المعلَمة، وأن العلاقة بين عدد المعالم وقيمة معامل الارتباط كانت متناقصة وتيرياً. وترى الباحثة أن النموذج ثلاثي المعلمة قدم تدريجاً آخر للقدرة معتمداً بذلك على احتساب أثر عامل التخمين الذي أهمله النموذجان أُحادي وثنائي المعلمة، اللذان قدما تدريجاً للقدرة يناظر تماماً تدريج العلامات الخام، لكن هذا التناظر التام لم يحدث بين العلامات الخام وعلامات القدرة في النموذج ثلاثي المعلمة.

النتائج المتعلقة بالدقة النسبية:

للبحث عن النموذج الذي يعطي معلومات أكبر وبالتالي يعطي تقديرات أكثر دقة للقدرة في التقدير، تم حساب الدقة النسبية، وذلك عن طريق الحصول على دالة المعلومات للاختبار عند كل نقطة على متصل السمة لكل نموذج. قامت الباحثة بعد ذلك بقسمة قيم دالة المعلومات الخاصة بالاختبار وفق النموذج ثنائي المعلَمة على قيم دالة المعلومات الخاصة بالاختبار وفق النموذج أُحادي المعلَمة، كذلك قامت بقسمة دالة المعلومات الخاصة بالاختبار وفق النموذج ثلاثي المعلَمة على دالة المعلومات الخاصة بالاختبار وفق النموذج أُحادي المعلَمة، يعرض الجدول 22 العلامة الخام على الاختبار والدقة النسبية لتقديرات القدرة.

جدول 22

العلامات الخام وقيم الدقة النسبية للفقرات المشتركة للاختبار وفق النماذج اللوجستية الثلاثة

العلامة الخام	الدقة النسبية م2/م1	الدقة النسبية م3/م1	العلامة الخام	الدقة النسبية م2/م1	الدقة النسبية م3/م1
9	0.973	0.008	28	1.047	1.943
10	0.992	0.007	29	1.047	2.110
11	0.987	0.010	30	1.042	2.312
12	1.015	0.057	31	1.035	2.479
13	1.015	0.058	32	1.040	2.617
14	1.026	0.104	33	1.025	2.810
15	1.029	0.170	34	1.023	2.966
16	1.036	0.154	35	1.027	3.134
17	1.041	0.334	36	0.998	3.275
18	1.037	0.419	37	0.999	3.420
19	1.047	0.596	38	0.980	3.484
20	1.049	0.700	39	0.975	3.668
21	1.052	0.854	40	0.964	3.669
22	1.053	0.981	41	0.995	3.699
23	1.055	1.176	42	0.952	3.706
24	1.055	1.245	43	0.940	3.561
25	1.055	1.420	44	0.946	3.679
26	1.055	1.574	45	0.913	3.487
27	1.053	1.752	46	0.909	4.141

م1: دالة معلومات الاختبار للنموذج اللوجستي أحادي المعلمة.

م2: دالة معلومات الاختبار للنموذج اللوجستي ثنائي المعلمة.

م3: دالة معلومات الاختبار للنموذج اللوجستي ثلاثي المعلمة.

يتبيّن من الجدول 22 أن هناك تقارباً في دقة التقدير وفي كمية المعلومات التي يقدمها الاختبار وفق النموذجين الأحادي والثنائي وعلى جميع مستويات القدرة تقريباً، إذ يلاحظ أن قيم الدقة تقارب الواحد صحيح حيث تراوحت هذه القيم بين 0.909 و 1.055. كانت قيم الدقة النسبية أقل من واحد تقريباً في المستويات الدنيا للقدرة والمستويات العليا أيضاً.

ولدى مقارنة كمية معلومات النموذج ثلاثي المعلَمة مع كمية معلومات النموذج أُحادي المعلَمة لم يقدم الأول معلومات تذكر بالمقارنة مع تلك التي قدمها الثاني عند المستويات الدنيا للقدرة، حيث تراوحت قيم الدقة النسبية للنموذج ثلاثي المعلمة بالنسبة إلى النموذج أُحادي المعلمة بين 0.008 و 0.419 في الثلث الأول من متصل القدرة لغاية θ_{30}، ثم بدأت قيم الدقة النسبية بالازدياد في الثلثين الأخيرين من متصل السمة، حيث تراوحت بين 0.596 و 4.141 ويلاحظ تفوق النموذج ثلاثي المعلمة على النموذج أُحادي المعلمة بدءاً بالعلامة الخام 23.

وكإجراء آخر يهدف إلى البحث عن النموذج الأفضل قامت الباحثة باستخدام محك أشار إليه كل من إمبرتسون وريس (Embretson & Reise, 2000) يعتمد على حساب مقدار التحسّن الذي يطرأ نتيجة الانتقال من نموذج إلى نموذج أكثر تعقيداً من خلال حساب الفرق بين قيمتي (-2LogLikelihood) لكل نموذج، باستخدام اختبار كاي تربيع. حيث يشير المقدار(-2LogLikelihood) إلى درجة انحراف البيانات عن النموذج، ويتم حساب مقدار التحسن الذي يضيفه الاختبار وفق النموذج ثنائي المعلمة عن الاختبار وفق النموذج أُحادي المعلمة بالمعادلة الآتية:

$$\chi^2 = -2LogLikelihood_{1pl} - (-2LogLikelihood_{2pl})$$
$$\chi^2 = 66847.4591 - 66667.4909$$
$$= 179.9682$$

بدرجات حرية 49، حيث تستخلص درجات الحرية من عدد المعالم الزيادة في النموذج الأكثر تعقيداً، إذ أضاف النموذج ثنائي المعلمة 49 معلماً زيادة عن معالم النموذج أُحادي المعلمة تعود لمعلم التمييز، وعند مقارنة قيم كاي تربيع المحسوبة مع قيم كاي تربيع الحرجة المستخرجة من الجدول عند درجات الحرية 49 نجدها 76.154 عند مستوى الدلالة ($\alpha = 0.01$).

كذلك قامت الباحثة بحساب مقدار التحسن الذي يضيفه الاختبار وفق النموذج ثلاثي المعلمة عن الاختبار وفق النموذج ثنائي المعلمة من خلال حساب قيمة كاي تربيع:

$$\chi^2 = -2\ \text{LogLikelihood}_{2pl} - (2-\ \text{LogLikelihood}_{3pl})$$

$$\chi^2 = 66667.4909 - 66531.9316$$

$$= 135.5593$$

أيضاً درجات الحرية تساوي 49، ولدى مقارنة قيمة كاي تربيع المحسوبة مع قيمة كاي تربيع الحرجة المستخرجة من الجدول وبدرجات حرية 49 نجد أن قيمة كاي تربيع المحسوبة أعلى.

كذلك قامت الباحثة بحساب مقدار التحسّن الذي يضيفه الاختبار وفق النموذج ثلاثي المعلمة على الاختبار وفق النموذج أُحادي المعلمة من خلال حساب قيمة كاي تربيع:

$$\chi^2 = -2\ \text{LogLikelihood}_{1pl} - (-\ 2\ \text{LogLikelihood}_{3pl})$$

$$\chi^2 = 66847.4591 - 66531.9316$$

$$= 315.5275$$

وبدرجات حرية 98 ولدى مقارنة قيمة كاي تربيع المحسوبة مع قيمة كاي تربيع الحرجة (135.807) يتبين أن قيمة كاي تربيع المحسوبة أكبر، أي أن النموذج ثلاثي المعلمة قدّم تقديرات أكثر دقة بالمقارنة مع النموذج أُحادي المعلمة وأيضاً ثنائي المعلمة.

النتائج المتعلقة بالفاعلية النسبية:

فيما يتعلق بالفاعلية النسبية قامت الباحثة باستلال اختبارات فرعية من الاختبار الكلي الممثل بالفقرات المشتركة (49 فقرة) لكل نموذج ومقارنة دالة المعلومات لكل اختبار فرعي مع دالة المعلومات للاختبار الكلي، حيث تمت مقارنة اختبار يمثل الفقرات السهلة وعدد فقراته 24 فقرة، وآخر يمثل الفقرات الصعبة أيضاً عدد فقراته 24 فقرة لكل نموذج مع كمية معلومات الاختبار. وتبين الجداول 23، 24، 25 الفاعلية النسبية للاختبارين (السهل والصعب) المستلين من الاختبار الكلي وفق النموذج الأحادي والثنائي والثلاثي على الترتيب.

جدول 23

العلامة الخام وتقديرات القدرة المتعادلة عند رتبٍ مئينية مختارة والخطأ المعياري في تقدير القدرة ودالة معلومات الاختبار والفاعلية النسبية لاختبارين فرعيين يحويان فقرات (سهلة، صعبة) مستلين من الاختبار الكلي الممثل بالفقرات المشتركة وفق النموذج اللوجستي أحادي المعلمة

المئين	العلامة الخام			القدرة			الخطأ المعياري			دالة المعلومات			الفاعلية النسبية	
	سهلة*	صعبة&	الكلي%	سهلة	صعبة	الكلي	سهلة	صعبة	الكلي	سهلة	صعبة	الكلي	س/ك	ص/ك
1	4	2	9	-2.655	-2.422	-1.955	0.742	0.946	0.496	1.816	1.117	4.061	0.447	0.275
2	5	3	10	-2.282	-1.840	-1.776	0.681	0.792	0.478	2.153	1.595	4.385	0.491	0.364
3	6	3	11	-1.959	-1.840	-1.609	0.640	0.792	0.462	2.445	1.595	4.686	0.522	0.340
4	6	3	12	-1.959	-1.840	-1.452	0.640	0.792	0.449	2.445	1.595	4.964	0.493	0.321
5	7	3	12	-1.671	-1.840	-1.452	0.610	0.792	0.449	2.691	1.595	4.964	0.542	0.321
10	8	4	13	-1.406	-1.406	-1.304	0.588	0.704	0.438	2.892	2.019	5.219	0.554	0.387
15	9	5	15	-1.167	-1.051	-1.026	0.573	0.647	0.420	3.048	2.392	5.661	0.539	0.423
20	10	5	16	-0.918	-1.051	-0.894	0.563	0.647	0.413	3.160	2.392	5.849	0.540	0.409
25	10	6	17	-0.918	-0.744	-0.766	0.563	0.607	0.408	3.160	2.714	6.015	0.525	0.451
30	11	6	18	-0.687	-0.744	-0.642	0.557	0.607	0.403	3.227	2.714	6.159	0.524	0.441
35	12	7	19	-0.458	-0.469	-0.520	0.555	0.579	0.399	3.249	2.984	6.282	0.517	0.475
40	12	7	20	-0.458	-0.469	-0.401	0.555	0.579	0.398	3.249	2.984	6.384	0.509	0.467
45	13	8	21	-0.230	-0.217	-0.283	0.557	0.559	0.393	3.227	3.205	6.465	0.499	0.496
46	13	8	21	-0.230	-0.217	-0.283	0.557	0.559	0.393	3.227	3.205	6.465	0.499	0.496
47	13	8	21	-0.230	-0.217	-0.283	0.557	0.559	0.393	3.227	3.205	6.465	0.499	0.496
48	13	8	21	-0.230	-0.217	-0.283	0.557	0.559	0.393	3.227	3.205	6.465	0.499	0.496
49	13	8	22	-0.230	-0.217	-0.167	0.557	0.559	0.391	3.227	3.205	6.525	0.495	0.491
50	14	8	22	0.002	-0.217	-0.167	0.563	0.559	0.391	3.160	3.205	6.525	0.484	0.491
51	14	8	22	0.002	-0.217	-0.167	0.563	0.559	0.391	3.160	3.205	6.525	0.484	0.491
52	14	8	22	0.002	-0.217	-0.167	0.563	0.559	0.391	3.160	3.205	6.525	0.484	0.491
53	14	8	23	0.002	-0.217	-0.051	0.563	0.559	0.390	3.160	3.205	6.565	0.461	0.488
54	14	9	23	0.002	0.021	-0.051	0.563	0.544	0.390	3.160	3.376	6.565	0.481	0.514
55	14	9	23	0.002	0.021	-0.051	0.563	0.544	0.390	3.160	3.376	6.565	0.481	0.514
60	15	9	24	0.240	0.021	0.064	0.573	0.544	0.390	3.049	3.376	6.583	0.463	0.513
65	16	10	26	0.490	0.249	0.294	0.588	0.535	0.390	2.892	3.497	6.559	0.441	0.533
70	16	11	27	0.490	0.470	0.410	0.588	0.529	0.392	2.892	3.569	6.515	0.444	0.548
75	17	11	29	0.755	0.470	0.645	0.610	0.529	0.398	2.691	3.569	6.367	0.423	0.561
80	18	13	31	1.043	0.907	0.887	0.640	0.529	0.404	2.445	3.567	6.136	0.398	0.581
85	19	14	33	1.365	1.129	1.140	0.681	0.535	0.414	2.153	3.493	5.821	0.370	0.600
90	20	16	36	1.739	1.595	1.551	0.742	0.559	0.439	1.816	3.199	5.169	0.350	0.616
95	21	18	40	2.196	2.124	2.207	0.836	0.608	0.498	1.432	2.707	4.036	0.355	0.671
96	21	19	41	2.196	2.431	2.403	0.836	0.647	0.521	1.432	2.388	3.691	0.388	0.647
97	22	20	42	2.609	2.787	2.619	0.999	0.705	0.549	1.002	2.015	3.320	0.302	0.607
98	22	21	43	2.609	3.222	2.881	0.999	0.793	0.585	1.002	1.591	2.925	0.343	0.544
99	23	22	44	3.808	3.605	3.141	1.380	0.947	0.632	0.525	1.115	2.505	0.209	0.445

* عدد الفقرات السهلة (24) فقرة.

& عدد الفقرات الصعبة (24) فقرة

% عدد الفقرات الكلية (49) فقرة

جدول 24

العلامة الخام وتقديرات القدرة المتعادلة عند رتبٍ مئينية مختارة والخطأ المعياري في تقدير القدرة ودالة معلومات الاختبار والفاعلية النسبية لاختبارين فرعيين يحويان فقرات (سهلة، وصعبة) مستلين من الاختبار الكلي الممثل بالفقرات المشتركة وفق النموذج اللوجستي ثنائي المعلَمة

المئيني	العلامة الخام			القدرة			الخطأ المعياري			دالة المعلومات			الفاعلية النسبية	
	سهلة *	صعبة &	كلي %	سهلة	صعبة	كلي	سهلة	صعبة	كلي	سهلة	صعبة	كلي	س/ك	ص/ك
1	4	2	9	-2.530	-2.524	-1.975	0.734	0.999	0.503	1.858	1.003	3.950	0.470	0.254
2	5	3	10	-2.168	-2.052	-1.772	0.671	0.856	0.479	2.218	1.364	4.351	0.510	0.313
3	6	3	11	-1.683	-2.052	-1.636	0.607	0.856	0.465	2.715	1.364	4.627	0.587	0.295
4	6	3	12	-1.683	-2.052	-1.430	0.607	0.856	0.445	2.715	1.364	5.040	0.539	0.271
5	7	3	12	-1.693	-2.052	-1.430	0.608	0.856	0.445	2.705	1.364	5.040	0.537	0.271
10	8	4	13	-1.144	-1.383	-1.301	0.558	0.698	0.435	3.210	2.053	5.296	0.606	0.388
15	9	5	15	-1.285	-0.863	-1.020	0.568	0.616	0.414	3.094	2.637	5.824	0.531	0.453
20	10	5	16	-0.944	-0.863	-0.885	0.546	0.616	0.406	3.349	2.637	6.057	0.553	0.435
25	10	6	17	-0.944	-0.765	-0.756	0.546	0.603	0.400	3.349	2.753	6.259	0.535	0.440
30	11	6	18	-0.407	-0.765	-0.667	0.533	0.603	0.396	3.523	2.753	6.388	0.552	0.440
35	12	7	19	-0.488	-0.429	-0.622	0.533	0.565	0.390	3.518	3.133	6.575	0.535	0.477
40	12	7	20	-0.488	-0.429	-0.404	0.533	0.565	0.386	3.523	3.133	6.699	0.526	0.468
45	13	8	21	-0.139	-0.378	-0.282	0.536	0.560	0.383	3.483	3.186	6.802	0.512	0.468
46	13	8	21	-0.139	-0.378	-0.282	0.536	0.560	0.383	3.483	3.186	6.802	0.512	0.468
47	13	8	21	-0.139	-0.378	-0.282	0.536	0.560	0.383	3.483	3.186	6.802	0.512	0.468
48	13	8	21	-0.139	-0.378	-0.282	0.536	0.560	0.383	3.483	3.186	6.802	0.512	0.468
49	13	8	22	-0.139	-0.378	-0.176	0.536	0.560	0.381	3.483	3.186	6.873	0.507	0.464
50	14	8	22	-0.171	-0.378	-0.176	0.535	0.560	0.381	3.492	3.186	6.873	0.508	0.464
51	14	8	22	-0.171	-0.378	-0.176	0.535	0.560	0.381	3.492	3.186	6.873	0.508	0.464
52	14	8	22	-0.171	-0.378	-0.176	0.535	0.560	0.381	3.492	3.186	6.873	0.508	0.464
53	14	8	23	-0.171	-0.378	-0.040	0.535	0.560	0.380	3.492	3.186	6.928	0.506	0.460
54	14	9	23	-0.171	-0.155	-0.040	0.535	0.542	0.380	3.492	3.400	6.928	0.508	0.491
55	14	9	23	-0.171	-0.155	-0.040	0.535	0.542	0.380	3.492	3.400	6.928	0.506	0.491
60	16	9	24	0.112	-0.155	0.036	0.545	0.542	0.380	3.369	3.400	6.943	0.485	0.490
65	16	10	26	0.448	0.266	0.271	0.566	0.521	0.380	3.122	3.664	6.917	0.451	0.533
70	16	11	27	0.448	0.408	0.390	0.566	0.517	0.382	3.122	3.735	6.864	0.455	0.544
75	17	11	29	0.699	0.408	0.636	0.589	0.517	0.387	2.887	3.735	6.669	0.433	0.560
80	18	12	31	0.864	0.622	0.898	0.609	0.516	0.397	2.698	3.763	6.353	0.424	0.592
85	19	14	33	1.408	1.023	1.149	0.684	0.523	0.409	2.134	3.659	5.988	0.358	0.613
90	20	16	36	1.677	1.633	1.577	0.714	0.561	0.439	1.960	3.180	5.180	0.378	0.614
95	21	18	40	2.039	2.271	2.233	0.811	0.636	0.507	1.519	2.475	3.891	0.390	0.636
96	21	19	41	2.039	2.255	2.347	0.811	0.633	0.522	1.519	2.493	3.674	0.414	0.679
97	22	20	42	2.931	3.042	2.634	1.066	0.777	0.562	0.880	1.655	3.161	0.278	0.524
98	22	21	43	2.931	3.649	2.882	1.066	0.933	0.603	0.880	1.150	2.750	0.320	0.418
99	23	22	44	3.832	4.000	3.142	1.433	999.000	0.650	0.487	0.000	2.370	0.205	0.000

* عدد الفقرات السهلة (24) فقرة.

& عدد الفقرات الصعبة (24) فقرة

% عدد الفقرات الكلية (49) فقرة

جدول 25

العلامة الخام وتقديرات القدرة المتعادلة عند رتبٍ مئينية مختارة والخطأ المعياري في تقدير القدرة ودالة معلومات الاختبار والفاعلية النسبية لاختبارين فرعيين يحويان فقرات (سهلة، صعبة) مستلين من الاختبار الكلي الممثل بالفقرات المشتركة وفق النموذج اللوجستي ثلاثي المعلَمة

المئيني	العلامة الخام			القدرة			الخطأ المعياري			دالة المعلومات			الفاعلية النسبية	
	سهلة*	صعبة&	الكلي%	سهلة	صعبة	الكلي	سهلة	صعبة	الكلي	سهلة	صعبة	الكلي	س/ك	ص/ك
1	4	3	9	-3.745	-2.216	-3.661	4.937	6.185	5.684	0.041	0.026	0.031	1.3	0.8
2	5	3	10	-3.803	-2.216	-3.813	5.156	6.185	5.817	0.038	0.026	0.030	1.3	0.9
3	5	3	11	-3.803	-2.216	-3.541	5.156	6.185	4.726	0.038	0.026	0.045	0.8	0.6
4	6	4	12	-2.094	-1.621	-2.207	1.549	3.477	1.878	0.417	0.083	0.284	1.5	0.3
5	6	4	12	-2.094	-1.621	-2.207	1.549	3.477	1.878	0.417	0.083	0.284	1.5	0.3
10	8	5	13	-1.438	-1.000	-2.008	1.011	1.896	1.825	0.978	0.278	0.300	3.3	0.9
15	8	5	15	-1.438	-1.000	-1.231	1.011	1.896	1.019	0.978	0.278	0.963	1.0	0.3
20	9	6	16	-0.940	-0.532	-1.277	0.752	1.211	1.054	1.767	0.682	0.900	2.0	0.8
25	10	6	17	-0.788	-0.532	-0.756	0.692	1.211	0.705	2.089	0.682	2.011	1.0	0.3
30	11	7	18	-0.519	-0.508	-0.615	0.603	1.174	0.622	2.750	0.726	2.583	1.1	0.3
35	11	7	19	-0.519	-0.508	-0.345	0.603	1.174	0.517	2.750	0.726	3.742	0.7	0.2
40	12	8	20	-0.143	-0.446	-0.215	0.515	1.105	0.473	3.771	0.819	4.466	0.8	0.2
45	13	8	21	-0.097	-0.446	-0.047	0.507	1.105	0.426	3.890	0.819	5.522	0.7	0.1
46	13	8	21	-0.097	-0.446	-0.047	0.507	1.105	0.426	3.890	0.819	5.522	0.7	0.1
47	13	8	21	-0.097	-0.446	-0.047	0.507	1.105	0.426	3.890	0.819	5.522	0.7	0.1
48	13	8	21	-0.097	-0.446	-0.047	0.507	1.105	0.426	3.890	0.819	5.522	0.7	0.1
49	13	8	22	-0.097	-0.446	0.073	0.507	1.105	0.395	3.890	0.819	6.402	0.6	0.1
50	13	8	22	-0.097	-0.446	0.073	0.507	1.105	0.395	3.890	0.819	6.402	0.6	0.1
51	13	8	22	-0.097	-0.446	0.073	0.507	1.105	0.395	3.890	0.819	6.402	0.6	0.1
52	14	9	22	0.012	0.308	0.073	0.490	0.561	0.395	4.160	3.183	6.402	0.6	0.5
53	14	9	23	0.012	0.308	0.230	0.490	0.561	0.360	4.160	3.183	7.720	0.5	0.4
54	14	9	23	0.012	0.308	0.230	0.490	0.561	0.360	4.160	3.183	7.720	0.5	0.4
55	14	9	23	0.012	0.308	0.230	0.490	0.561	0.360	4.160	3.183	7.720	0.5	0.4
60	15	9	24	0.410	0.308	0.282	0.455	0.561	0.349	4.630	3.183	8.197	0.6	0.4
65	16	10	26	0.548	0.294	0.496	0.452	0.587	0.311	4.904	3.111	10.325	0.5	0.3
70	16	11	27	0.548	0.794	0.597	0.452	0.403	0.296	4.904	6.165	11.418	0.4	0.5
75	17	12	29	0.788	0.912	0.781	0.455	0.379	0.273	4.824	6.952	13.437	0.4	0.5
80	18	13	31	1.006	1.018	0.963	0.469	0.362	0.256	4.553	7.614	15.214	0.3	0.5
85	19	14	33	1.377	1.110	1.113	0.511	0.351	0.247	3.828	8.131	16.357	0.2	0.5
90	20	16	36	1.493	1.549	1.363	0.530	0.329	0.243	3.564	9.250	16.994	0.2	0.5
95	21	19	40	1.607	2.145	1.746	0.550	0.380	0.260	3.306	6.934	14.808	0.2	0.5
96	21	19	41	1.607	2.145	1.851	0.550	0.380	0.271	3.306	6.934	13.651	0.2	0.5
97	22	20	42	2.278	2.142	1.985	0.716	0.379	0.285	1.948	6.953	12.303	0.2	0.8
98	22	21	43	2.278	2.487	2.137	0.716	0.449	0.310	1.948	4.963	10.416	0.2	0.5
99	23	22	44	3.172	2.510	2.235	1.081	0.455	0.329	0.858	4.836	9.217	0.1	0.5

* عدد الفقرات السهلة (24) فقرة.

& عدد الفقرات الصعبة (24) فقرة

% عدد الفقرات الكلية (49) فقرة

تُظهر البيانات الخاصة بالنموذج أحادي المعلَمة أن الاختبار الذي يمثل الفقرات السهلة (24 فقرة) قدّم تقريباً نصف كمية المعلومات التي قدمها الاختبار الكلي (49 فقرة) لغاية م$_{55}$، حيث تراوحت قيم الفاعلية النسبية بين 0.481 و 0.554 ما عدا قيمة الفاعلية عند م$_{1}$ حيث كانت 0.447، ثم بدأت قيم الفاعلية للفقرات السهلة بالانخفاض التدريجي حتى وصلت إلى 0.209 عند م$_{99}$. ولدى مقارنة قيم الفاعلية النسبية للفقرات السهلة عند المستويات الدنيا للقدرة مع المستويات العليا نجد أن الفقرات السهلة قدمت معلومات في الثلث الأول من متصل القدرة أفضل من المعلومات التي قدمتها في الثلث الأخير من متصل القدرة.

أما بخصوص الفاعلية النسبية للاختبار الذي يمثل الفقرات الصعبة فقد تراوحت بين 0.257 و 0.451 لغاية م$_{30}$، في حين تراوحت بين 0.544 و 0.671 بين م$_{70}$ و م$_{98}$، وانخفضت قيم الفاعلية إلى 0.445 عند م$_{99}$. بشكل عام قدم الاختبار الذي يمثل الفقرات الصعبة معلومات في الثلث الأخير من متصل القدرة أفضل من المعلومات التي قدمها في الثلث الأول.

يظهر الجدول 24 قيماً للفاعلية النسبية تتراوح بين 0.470 و0.606 لغاية م$_{60}$ لكنها اتسمت بالانخفاض التدريجي، فيما تراوحت هذه القيم بين 0.205 و 0.455 في الثلث الأخير تقريباً لمتصل القدرة، أي قدّم الاختبار الذي يمثل الفقرات السهلة معلومات في بداية متصل السمة أكبر من تلك التي في نهايته، وبدأت القيم بالانخفاض التدريجي حتى وصلت إلى 0.205 عند م$_{99}$.

وفيما يتعلق بالفاعلية النسبية للاختبار الذي يمثل الفقرات الصعبة أظهرت النتائج قيماً للفاعلية في الثلث الأول لمتصل القدرة أقل من تلك القيم التي ظهرت في الثلث الأخير تقريباً. حيث تراوحت هذه القيم بين 0.254 و 0.453 لغاية م$_{30}$، فيما تراوحت بين 0.418 و 0.679 في الثلث الأخير تقريباً.

يتبيّن من الجدول 25 أن الفاعلية النسبية للاختبار وفق النموذج ثلاثي المعلَمة كانت أعلى لدى الفقرات السهلة عند مستويات القدرة الدنيا، وتفوقت على الاختبار الكلي لغاية م30، باستثناء م3، بدأت بعد ذلك قيم الفاعلية بالانخفاض التدريجي حيث أشارت النتائج إلى أن قيم الفاعلية في الثلث الأول لمتصل القدرة تراوحت بين 0.840 و 3.256 في حين تراوحت هذه القيم بين 0.093 و 0.429 في الثلث الأخير لمتصل القدرة.

وفيما يتعلق بقيم الفاعلية النسبية للاختبار الذي يمثل الفقرات الصعبة يشير الجدول إلى أن قيم الفاعلية بشكل عام في الثلث الأخير لمتصل القدرة أشارت إلى أن الاختبار الذي يمثل الفقرات الصعبة قدّم دقة تعادل نصف دقة الاختبار الكلي، كما تشير النتائج إلى أن قيم الفاعلية النسبية في منطقة الوسط من متصل القدرة أبدت انخفاضاً واضحاً حيث تراوحت الفاعلية بين 0.128 و 0.540.

النتائج المتعلقة بالتحقق من ميزات (خصائص) النموذج:

1. النتائج المتعلقة بالتحقق من خاصية اللاتغير لتقديرات معالم الفقرة:

تم حساب معامل ارتباط سبيرمان (الرتب) لمعالم الفقرات للمجموعات متساوية العدد التي تم أخذها من العينة الكلية وهي: ذكور 1، ذكور 2، إناث 1، وإناث 2، حيث بلغ عدد أفراد كل مجموعة من المجموعات الأربعة 244 فرداً. يبيّن الجدول 26 قيم معاملات الارتباط للمعالم الثلاثة للاختبار وفق النماذج اللوجستية الثلاثة لجميع المجموعات.

جدول 26

معاملات ارتباط سبيرمان لمعالم فقرات الاختبار وفق النماذج اللوجستية الثلاثة للمجموعات المختلفة

	النموذج					
	أحادي	ثنائي		ثلاثي		
	معلم الصعوبة	معلم الصعوبة	معلم التمييز	معلم الصعوبة	معلم التمييز	معلم التخمين
م1*م2	0.96	0.95	1.00	0.94	1.00	1.00
م1*م3	0.86	0.85	1.00	0.81	1.00	1.00
م1*م4	0.85	0.84	1.00	0.73	1.00	1.00
م2*م3	0.86	0.86	1.00	0.79	1.00	1.00
م2*م4	0.84	0.82	1.00	0.72	1.00	1.00
م3*م4	0.93	0.92	1.00	0.87	1.00	1.00

م1: مجموعة الإناث الأولى

م2: مجموعة الإناث الثانية

م3: مجموعة الذكور الأولى

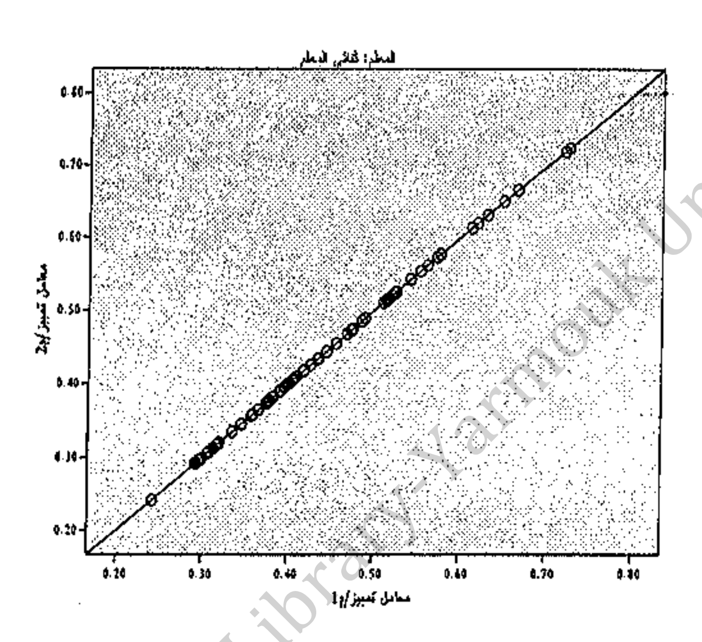
م4: مجموعة الذكور الثانية

يتبيّن من الجدول أن معاملات الارتباط لمعلَم الصعوبة تراوحت بين 0.72 لمجموعتي إناث 2، وذكور 2 في النموذج ثلاثي المعلَمة و0.96 لمجموعتي إناث 1، إناث 2 في النموذج أُحادي المعلَمة، كما اتسمت معاملات الارتباط لمعلم الصعوبة في النموذج ثلاثي المعلمة نوعاً ما بأنها الأقل بين النماذج الثلاثة، وتقاربت المعاملات المتناظرة للنموذجين الأُحادي والثنائي.

كما بيّن الجدول أن جميع قيم معاملات الارتباط الخاصة بمعلَم التمييز للنموذجين تساوي واحداً. وأظهر الجدول أن جميع معاملات الارتباط لمعلَم التخمين ولجميع المجموعات يساوي واحداً.

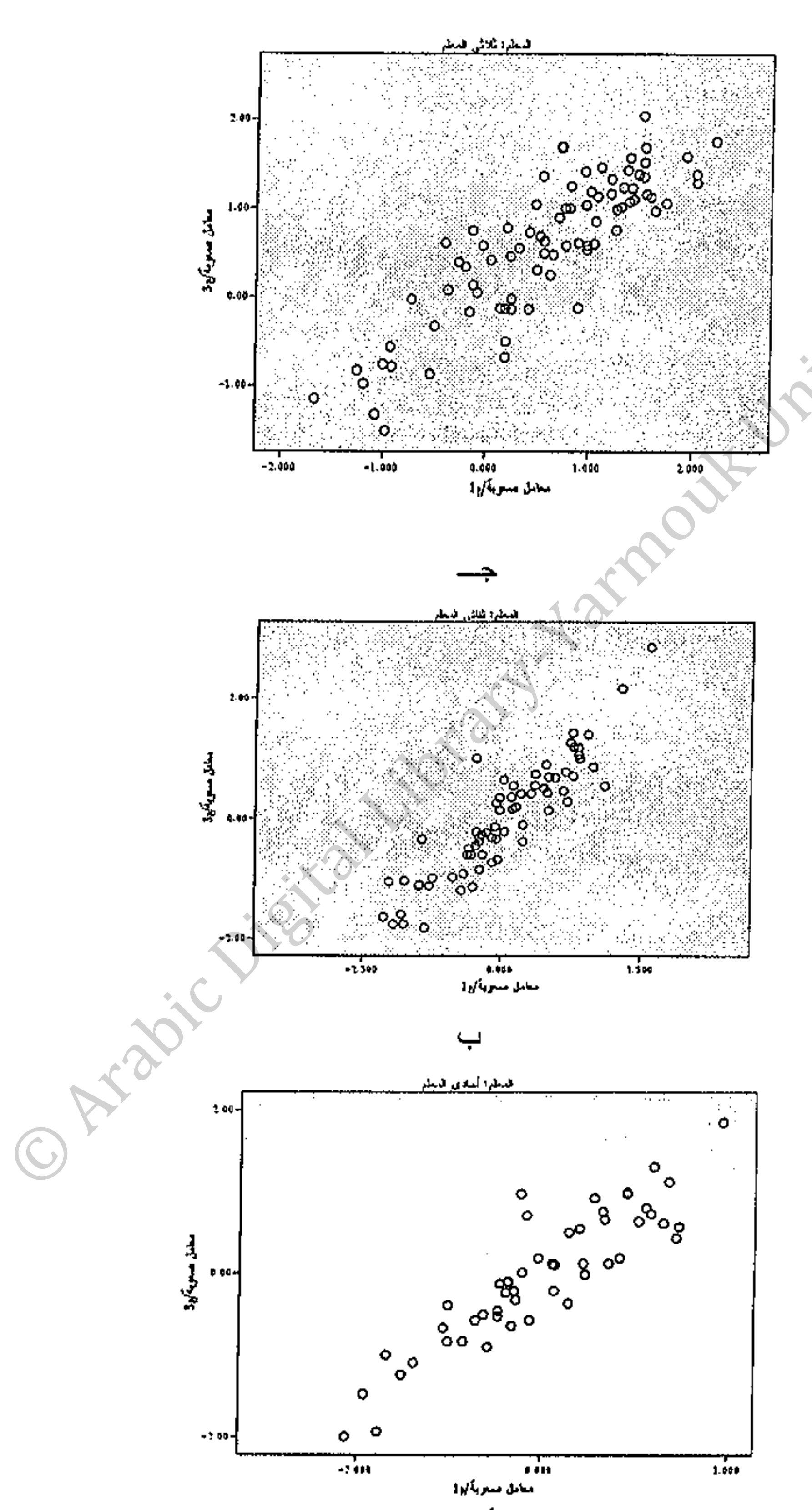
وتم رسم شكل الانتشار (Scatter) لمعالم الفقرات لكل مجموعتين، تبيّن الأشكال

7-5 التمثيل البياني لهذه المعالم.

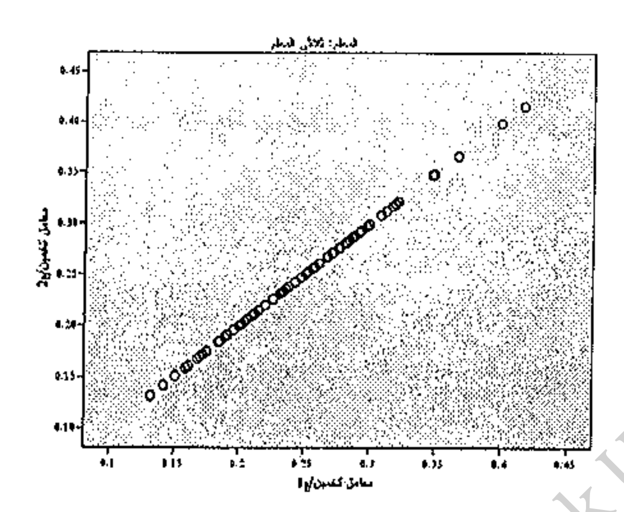


الشكل 5: شكل الانتشار لمعلّم التمييز للمجموعتين إناث1، إناث 2، لفقرات الاختبار وفق

النموذج اللوجستي ثنائي المعلّمة



الشكل 6: شكل الانتشار لمعلَم الصعوبة للمجموعتين إناث 1، ذكور 1، لفقرات الاختبار وفق النموذج أحادي المعلَمة في أ، ثنائي المعلَمة في ب، ثلاثي المعلَمة في ج.



الشكل 7: شكل الانتشار لمعلَم التخمين لفقرات الاختبار وفق النموذج ثلاثي المعلَمة للمجموعتين إناث 1، إناث 2.

كما قامت الباحثة بحساب المتوسطات الحسابية والانحرافات المعيارية لمعالم الفقرة لكل نموذج، تبيّن الجداول 27–29، هذه النتائج.

جدول 27

المتوسطات الحسابية والانحرافات المعيارية لمعلم الصعوبة لجميع المجموعات للاختبار وفق النماذج اللوجستية الثلاثة

المجموعة	النموذج					
	أحادي المعلمة		ثنائي المعلمة		ثلاثي المعلمة	
	المتوسط	الانحراف	المتوسط	الانحراف	المتوسط	الانحراف
مجموعة الإناث الأولى	0.226	0.92	0.246	0.95	0.894	0.69
مجموعة الإناث الثانية	0.221	0.82	0.243	0.87	0.837	0.61
مجموعة الذكور الثالثة	0.028	0.73	0.066	0.74	0.797	0.52
مجموعة الذكور الرابعة	0.065	0.73	0.095	0.75	0.757	0.52

يظهر الجدول تقارب كبير لقيم المتوسطات الحسابية لمجموعتي الإناث الأولى والثانية وفق النماذج اللوجستية الثلاثة، وتقارب لقيم المتوسطات الحسابية لمجموعتي الذكور الأولى والثانية بشكل معتدل مع ابتعاد متوسطات الذكور عن متوسطات الإناث بشكل عام.

جدول 28

المتوسطات الحسابية والانحرافات المعيارية لمعلَم التمييز لجميع المجموعات للاختبار وفق النموذجين ثنائي وثلاثي المعلَمة

المجموعة	النموذج			
	ثنائي المعلمة		ثلاثي المعلمة	
	المتوسط	الانحراف	المتوسط	الانحراف
مجموعة الإناث الأولى	0.412	0.08	0.811	0.23
مجموعة الإناث الثانية	0.412	0.08	0.811	0.23
مجموعة الذكور الأولى	0.412	0.08	0.811	0.23
مجموعة الذكور الثانية	0.412	0.08	0.811	0.23

جدول 29

المتوسطات الحسابية والانحرافات المعيارية لمعلَم التخمين لجميع المجموعات للاختبار وفق النموذج ثلاثي المعلَمة

المجموعة	النموذج	
	ثلاثي المعلمة	
	المتوسط	الانحراف
مجموعة الإناث الأولى	0.236	0.07
مجموعة الإناث الثانية	0.236	0.07
مجموعة الذكور الأولى	0.236	0.07
مجموعة الذكور الثانية	0.236	0.07

يظهر الجدولان 28، 29 قيماً متساوية للمتوسطات الحسابية والانحرافات المعيارية للمعالم الخاصة بالمجموعات الأربع لمعلمي التمييز والتخمين.

استخدمت الباحثة تحليل التباين الأُحادي لحساب دلالة الفرق بين المتوسطين لمعلم الصعوبة، يبيّن الجدول 30 هذه النتائج للنماذج الثلاثة.

جدول 30

تحليل التباين الأُحادي للمتوسطات الحسابية لمعلَم الصعوبة لجميع المجموعات للاختبار وفق النماذج اللوجستية الثلاثة

النموذج	مصدر التباين	مجموعة المربعات	درجة الحرية	متوسط المربعات	قيمة ف المحسوبة	الدلالة الإحصائية	الدلالة العملية
	بين المجموعات	1.565	3	0.522	0.803	0.494	0.012
	داخل المجموعات	124.706	192	0.650			
أحادي المعلمة	الكلي	126.271	195				
	بين المجموعات	1.345	3	0.448	0.647	0.586	0.01
	داخل المجموعات	133.137	192	0.693			
ثنائي المعلمة	الكلي	134.482	195				
	بين المجموعات	0.502	3	0.167	0.480	0.696	0.007
	داخل المجموعات	66.949	192	0.349			
ثلاثي المعلمة	الكلي	67.452	195				

يتبيّن من الجدول 30 أن الفروق بين المتوسطات الحسابية التقديرة لمعلم الصعوبة غير دالة إحصائياً ($\alpha = 0.01$).

2. النتائج المتعلقة بالتحقق من خاصية اللاتغير لتقديرات القدرة:

قامت الباحثة بإيجاد المتوسطات الحسابية والانحرافات المعيارية لقدرات الأفراد على كل من الفقرات الصعبة والسهلة. يبيّن الجدول 31 هذه المتوسطات والانحرافات المعيارية ودلالة الفروق بين المتوسطات الحسابية.

جدول 31

نتائج اختبار ت للعينات المترابطة لمتوسطات قدرات الأفراد على الاختبارين (السهل والصعب)

النموذج	قدرة الفرد على الفقرات	العدد	المتوسط	الانحراف	معامل ارتباط بيرسون	قيمة ت المحسوبة	درجة الحرية	الدلالة الإحصائية
	السهلة	1086	0.035	1.25	0.702	2.053	1085	0.040
أحادي المعلمة	الصعبة	1086	0.024-	1.21				
	السهلة	1086	0.033	1.25	0.699	2.483	1085	0.013
ثنائي المعلمة	الصعبة	1086	0.040-	1.22				
	السهلة	1086	0.061-	1.36	0.562	1.389	1085	0.165
ثلاثي المعلمة	الصعبة	1086	0.114-	1.36				

يتبيّن من الجدول 31 أن المتوسطات الحسابية للفقرات الصعبة والسهلة كان – 0.024 و 0.035، على الترتيب للاختبار وفق النموذج أُحادي المعلَمة، ولم يكن الفرق دالاً إحصائياً (α = 0.01). ويظهر الجدول أن الانحراف المعياري لقدرات الأفراد على كل من الفقرات الصعبة والسهلة كان على الترتيب 1.21 و 1.25.

وأما ما يتعلق بالمتوسطات الحسابية لقدرات الأفراد على كل من الفقرات الصعبة والسهلة لنموذج ثنائي المعلَمة فقد كانت على الترتيب –0.04 و0.033 ولم يكن الفرق بين المتوسطين دالاً إحصائياً (α = 0.01)، وقد كان الانحراف المعياري لتقديرات قدرات الأفراد على الفقرات الصعبة والسهلة على الترتيب 1.22 و 1.25.

ويظهر الجدول 31 أن المتوسطين الحسابيين لتقديرات القدرة المقابلين للفقرات الصعبة والسهلة لفقرات النموذج ثلاثي المعلمة كانا على التوالي –0.114 و –0.061، ولم يكن الفرق دالاً إحصائياً (α = 0.01)، وكانت قيمة الانحرافين المعياريين لتقديرات القدرة للفقرات الصعبة والسهلة متساوية 1.36.

كما قامت الباحثة بحساب معامل ارتباط سبيرمان بين كل تقديرين لمعلَم القدرة في كل مجموعة، يبيّن الجدول 32 معاملات ارتباط سبيرمان للاختبار وفق النماذج الثلاثة.

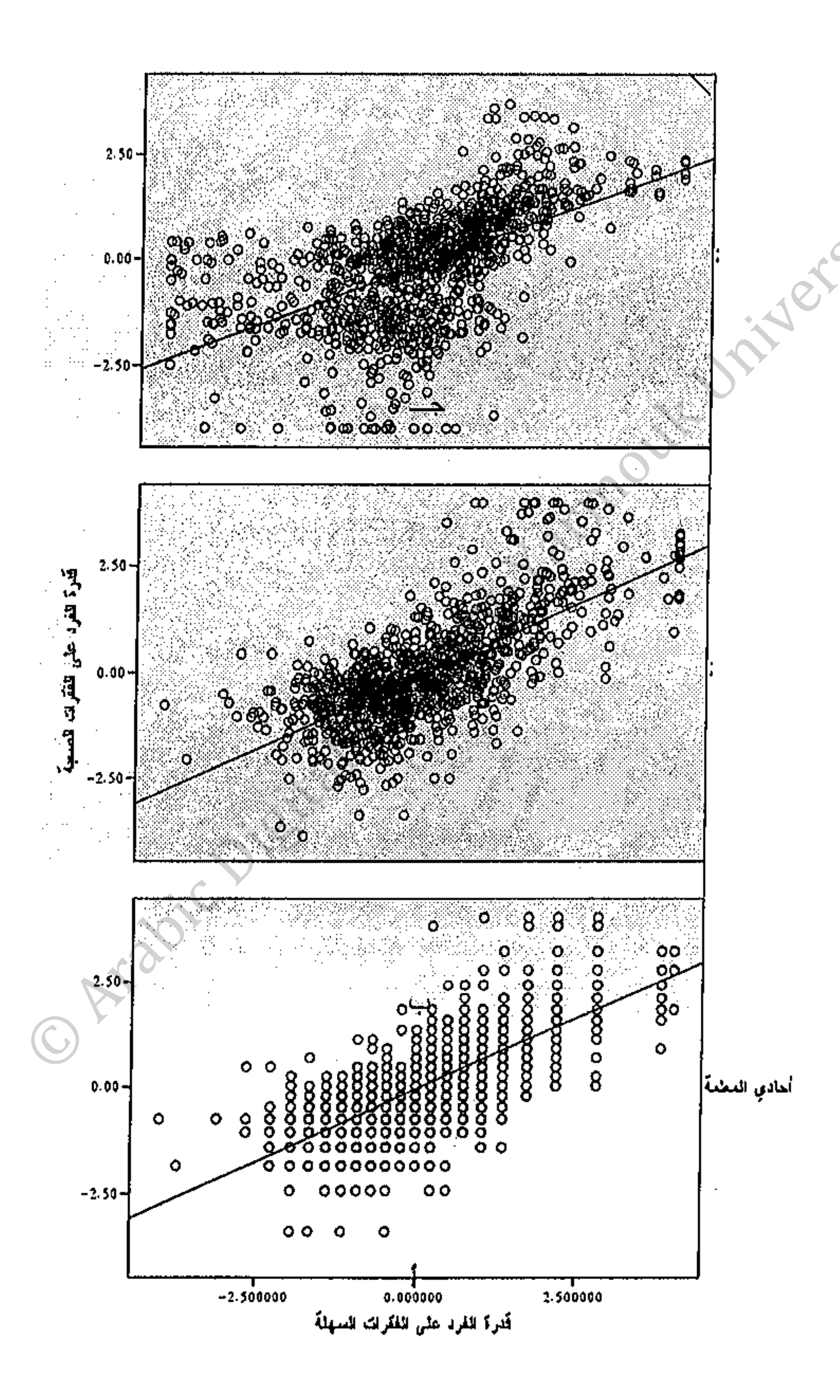
جدول 32

معاملات ارتباط سبيرمان بين تقديرات القدرة للأفراد على الاختبارين السهل والصعب

النموذج	معامل الارتباط
أحادي المعلمة	0.68
ثنائي المعلمة	0.67
ثلاثي المعلمة	0.65

يشير الجدول 32 إلى اتساق النتائج تقريباً للنماذج اللوجستية الثلاثة، إذ أبدت جميـــع القيم علاقة إيجابية معتدلة.

كما قامت الباحثة برسم شكل الانتشار لكل تقديرين في كل مجموعة. يبيّن الـشكل 8 شكل الانتشار لكل نموذج.



الشكل 8: شكل الانتشار لقدرات الأفراد على مجموعتي الفقرات السهلة والصعبة للاختبار وفق النموذج أُحادي المعلّمة في أ. ووفق النموذج ثنائي المعلم في ب، ووفق النموذج ثلاثي المعلمة في ج.

البيانات الوصفية للاختبار الكلي (89 فقرة) والاختبار بصورته النهائية (82 فقرة):

قامت الباحثة بحساب المتوسط الحسابي للعلامات على الاختبار الكلي (89 فقرة)، لجميع البيانات وللعينة الكلية (1174)، بالإضافة إلى حساب كافة المؤشرات الإحصائية الوصفية اللازمة، يبيّن الجدولان 33-34 هذه الإحصائيات.

جدول 33

المؤشرات الإحصائية الوصفية للعلامات الخام على الاختبار الكلي (89 فقرة)

الإحصائي	المستوى التعليمي		
	المرحلة الأساسية العليا	المرحلة الثانوية	المرحلة الجامعية
العدد	444	610	120
المتوسط	38.82	46.74	55.02
الخطأ المعياري للمتوسط	0.59	0.56	1.36
الوسيط	37	45	55
المنوال	33	45	*38
الانحراف المعياري	12.40	13.83	14.85
التباين	153.76	191.32	220.54
الالتواء	0.87	0.33	0.04-
الخطأ المعياري للالتواء	0.12	0.10	0.22
التفلطح	0.89	0.38-	0.87-
الخطأ المعياري للتفلطح	0.23	0.20	0.44
المدى	66	75	62
أدنى علامة مشاهدة	14	7	22
أعلى علامة مشاهدة	80	82	84

* التوزيع عديد المنوال، تم اختيار أصغر منوال.

جدول 34

المتوسطات الحسابية والانحرافات المعيارية لمجموعات الدراسة المختلفة وفق المستوى التعليمي والجنس لبيانات الاختبار الكلي

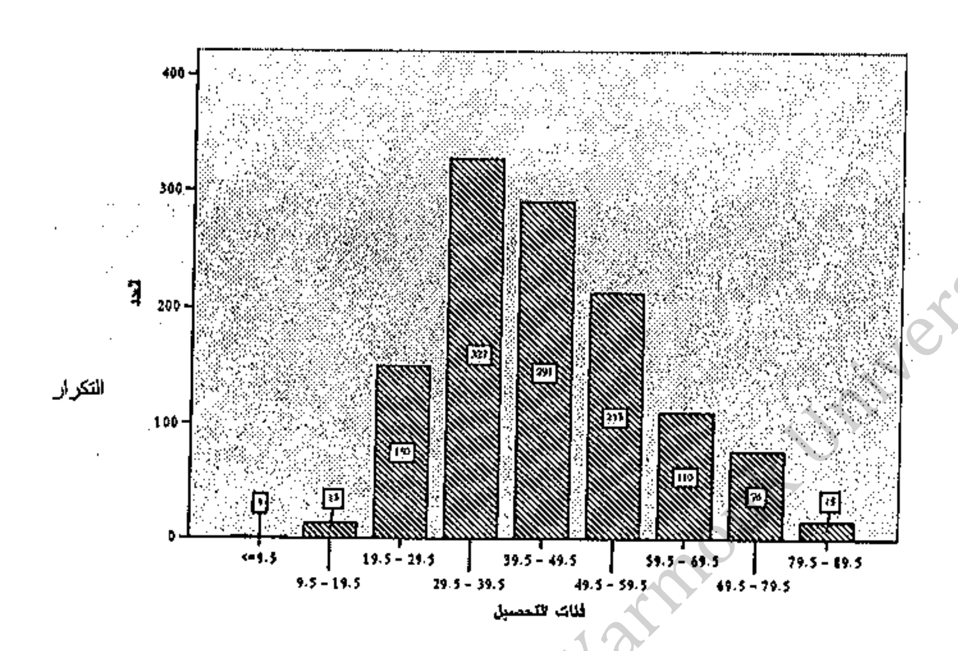
المستوى التعليمي	الجنس						الكلي		
	أنثى			ذكر					
	المتوسط	الانحراف	العدد	المتوسط	الانحراف	العدد	المتوسط	الانحراف	العدد
المرحلة الأساسية العليا	40.804	11.865	265	35.883	12.623	179	38.820	12.400	444
المرحلة الثانوية	47.490	12.506	290	46.069	14.919	320	46.744	13.832	610
المرحلة الجامعية	50.525	15.475	59	59.361	12.919	61	55.017	14.851	120
الكلي	**44.896**	**13.057**	**614**	**44.261**	**15.649**	**560**	**44.593**	**14.350**	**1174**

قامت الباحثة بتكوين جدول تكراري لعلامات الأفراد على الاختبار الكلي، ومثلت هذا الجدول بالأعمدة البيانية، يبيّن الجدول 35 التوزيع التكراري لبيانات الأفراد على الاختبار الكلي، في حين يمثل الشكل 9 التمثيل البياني لهذا الجدول.

جدول 35

التوزيع التكراري للعلامات الخام على الاختبار الكلي (89 فقرة)

الفئة	التكرار	التكرار النسبي	النسبة التراكمية
10 >	1	0.0009	0.0009
19-10	13	0.0111	0.012
29-20	130	0.1105	0.1225
39-30	327	0.2781	0.4006
49-40	291	0.2470	0.6476
59-50	213	0.1811	0.8287
69-60	110	0.0935	0.9222
79-70	76	0.0646	0.9826
89-80	15	0.0128	1



الشكل 9: التمثيل بالأعمدة البيانية للجدول التكراري للعلامات الخام على الاختبار الكلي

وقامت الباحثة أيضاً بحساب المؤشرات الإحصائية الوصفية أيضاً للعلامات الخام الخاصة بالاختبار بصورته النهائية (82 فقرة) والمطابقة للنموذج ثلاثي المعلمة. يبين الجدولان 36-37 المؤشرات الإحصائية الوصفية للعلامات الخام للاختبار بصورته النهائية (82 فقرة).

جدول 36

المؤشرات الإحصائية الوصفية للبيانات الخام على الاختبار بصورته النهائية

الإحصائي	المستوى التعليمي		
	المرحلة الأساسية العليا	المرحلة الثانوية	المرحلة الجامعية
العدد	410	566	102
أدنى علامة	17	17	19
المتوسط	35.910	42.989	52.284
الوسيط	33	42	52.5
المنوال	30	30	*67
أعلى علامة	73	76	77
الانحراف المعياري	11.45	12.92	14.04
الالتواء	1.03	0.40	-0.27
التفلطح	1.06	-0.55	-0.85

*التوزيع عديد المنوال، تم اختيار أصغر منوال.

يتبين من الجدول 36 أن الاختبار ملتوٍ للمجموعات الثلاث، ولكن الالتواء سالباً لطلبة المرحلة الجامعية.

جدول 37

المتوسطات الحسابية والانحرافات المعيارية لمجموعات الدراسة المختلفة وفق المستوى التعليمي والجنس لبيانات الاختبار بصورته النهائية (82 فقرة)

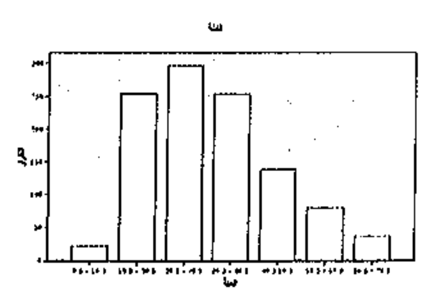
المستوى التعليمي	الجنس: أنثى			الجنس: ذكر			الكلي		
	المتوسط الحسابي	الانحراف المعياري	العدد	المتوسط	الانحراف المعياري	العدد	المتوسط الحسابي	الانحراف المعياري	العدد
المرحلة الجامعية	47.882	14.93	51	56.686	11.64	51	52.284	14.04	102
المرحلة الثانوية	43.241	11.85	282	42.739	13.91	284	42.989	12.92	566
المرحلة الأساسية العليا	37.085	11.08	260	33.873	11.82	150	35.910	11.45	410
الكلي	40.941	12.35	593	41.464	14.61	485	41.176	13.41	1078

كما قامت الباحثة بتكوين جدول تكراري لعلامات الأفراد على الاختبار بصورته النهائية، ومثلت هذا الجدول بالأعمدة البيانية. ويبيّن الجدول 38 التوزيعات التكرارية لبيانات الأفراد على الاختبار بصورته النهائية، في حين يبين الشكل 10 التمثيل البياني لهذا الجدول.

جدول 38

التوزيع التكراري للعلامات الخام على الاختبار بصورته النهائية (82 فقرة)

الفئة	التكرار	التكرار النسبي	النسبة التراكمية%
من 10 - 19	22	0.020	0.020
من 20 - 29	254	0.236	0.256
من 30 - 39	296	0.275	0.531
من 40 - 49	253	0.235	0.765
من 50 - 59	138	0.128	0.893
من 60 - 69	79	0.073	0.967
من 70 - 79	36	0.033	1.000
الكلي	1078	1.000	



الشكل 10: التمثيل بالأعمدة البيانية للجدول التكراري للعلامات الخام على الاختبار بصورته النهائية.

يتبيّن من الجدول والشكل أن التكرار الأكبر كان لفئة العلامات 30 - 39.

وقامت الباحثة بحساب الرتب المئينية للمجموعات المعيارية الآتية: المرحلة الأساسية العليا والمرحلة الثانوية والمرحلة الجامعية، لبيانات الاختبار الكلي. يبيّن الجدول 39 هذه النتائج.

جدول 39

الرتب المئينية للمجموعات المعيارية المختلفة: المرحلة الأساسية العليا والمرحلة الثانوية والمرحلة الجامعية لبيانات الاختبار الكلي (89 فقرة)

العلامة الخام	الرتبة المئينية		
	المرحلة الأساسية العليا	المرحلة الثانوية	المرحلة الجامعية
14	0.23	0.33	
16	0.45	0.41	
17	0.68	0.49	
18	0.90	0.74	
19	1.58	0.88	
20	2.70	1.01	
21	3.72	1.15	
22	5.29	1.46	0.83
23	7.32	2.13	1.67
24	9.35	3.03	1.95
25	11.37	3.61	2.22
26	13.51	4.18	2.50
27	16.67	5.16	2.92
28	19.03	6.89	3.33
29	21.28	8.69	4.17
30	24.77	9.92	5.00
31	28.15	11.64	5.42
32	31.76	14.51	5.83
33	35.70	17.13	6.25
34	39.75	19.67	6.67
35	43.81	22.46	7.50
36	47.75	25.08	8.33
37	51.13	27.87	10.42
38	53.49	30.82	14.58
39	55.97	33.62	18.75
40	58.67	35.33	20.83
41	61.71	36.64	21.46
42	65.32	38.77	22.08
43	68.24	42.13	24.58
44	70.95	45.00	27.50
45	73.65	48.20	29.17
46	75.68	51.80	30.00
47	77.59	54.34	32.08
48	79.95	56.80	34.17
49	81.98	59.34	35.42
50	83.56	61.89	38.33
51	85.47	64.26	42.08
52	87.16	66.97	45.42
53	88.40	69.26	47.92
54	89.53	70.57	49.58
55	90.20	72.79	51.67
56	90.99	75.57	53.75
57	91.89	77.79	55.42
58	92.57	79.43	58.33
59	93.24	80.82	62.50
60	93.81	82.13	65.42
61	94.59	83.36	66.05
62	95.27	84.51	66.67
63	95.72	85.66	67.50
64	96.28	86.89	68.33
65	96.73	88.36	70.42
66	96.84	89.51	72.92
67	96.96	90.08	74.17
68	97.07	91.23	75.83
69	97.18	92.70	77.50
70	97.27	94.10	78.75
71	97.36	95.16	80.00
72	97.48	95.74	83.33
73	97.55	96.23	87.08
74	97.64	96.56	88.55
75	97.87	96.89	90.00
76	98.09	97.22	90.83
77	98.87	97.54	91.67
78	99.66	98.20	94.58
79	99.83	98.69	97.92
80	100.00	99.18	98.34
81		99.59	98.75
82		100.00	99.17
83			99.59
84			100.00

كما قامت الباحثة بحساب الرتب المئينية لنفس المجموعات المعيارية السابقة لبيانات الاختبار بصورته النهائية (82 فقرة)، يمثل الجدول 40 هذه النتائج.

جدول 40

الرتب المئينية للمجموعات المعيارية المختلفة: المرحلة الأساسية العليا والمرحلة الثانوية والمرحلة الجامعية لبيانات الاختبار بصورته النهائية (82 فقرة)

العلامة الخام	المستوى التعليمي		
	الرتبة المئينية		
	المرحلة الأساسية العليا	المرحلة الثانوية	المرحلة الجامعية
17	0.24	0.18	
18	0.61	0.36	
19	1.22	0.53	0.98
20	2.68	0.71	1.18
21	5.00	1.50	1.37
22	7.32	2.39	1.57
23	9.63	3.45	1.76
24	12.44	4.68	1.96
25	15.73	6.01	3.43
26	19.27	7.33	4.17
27	22.32	9.63	4.90
28	25.98	12.01	5.23
29	29.88	14.22	5.56
30	34.88	17.84	5.88
31	40.49	21.29	6.86
32	44.39	23.50	7.84
33	48.54	26.68	11.27
34	51.83	30.21	12.99
35	54.51	32.33	14.71
36	57.44	34.19	18.14
37	60.61	36.57	20.59
38	63.78	38.52	21.33
39	66.83	40.84	22.06
40	70.12	43.55	24.02
41	72.56	47.08	24.76
42	74.51	51.24	25.49
43	76.46	54.33	25.98
44	79.15	56.80	26.47
45	81.83	59.28	27.94
46	83.41	61.84	30.88
47	85.37	64.58	34.80
48	87.20	67.05	38.73
49	88.54	69.61	41.18
50	89.63	71.82	43.63
51	90.24	73.59	46.57
52	90.85	75.62	49.02
53	91.59	77.74	51.47
54	92.32	79.59	53.43
55	92.93	81.01	56.86
56	93.78	82.16	59.80
57	94.51	83.57	60.78
58	95.12	85.07	61.52
59	95.61	86.66	62.25
60	95.98	87.99	65.69
61	96.46	88.78	69.12
62	96.58	89.84	70.59
63	96.71	91.87	71.57
64	96.83	93.29	73.53
65	96.95	93.99	75.98
66	97.11	95.05	77.94
67	97.20	95.85	81.86
68	97.44	96.29	86.76
69	97.80	96.64	88.40
70	98.05	97.08	90.03
71	98.54	97.35	91.67
72	99.15	97.88	95.59
73	99.76	98.76	97.55
74		99.38	98.04
75		99.73	98.53
76		100.00	99.02
77			100.00

الفصل الخامس

مناقشة النتائج والتوصيات

يتناول هذا الفصل خلاصة لنتائج الدراسة، ومناقشتها، وتفسيرها، كما يتناول عرضاً للتوصيات والمقترحات التي توصي بها الباحثة، حيث كان الغرض من هذه الدراسة بناء اختبار في مهارات التفكير الناقد وفق نظرية استجابة الفقرة.

- مناقشة النتائج المتعلقة بالتحقق من افتراضات نماذج نظرية استجابة الفقرة:

أظهرت نتائج التحليل العاملي من الرتبة الثانية والمتبوع بتدوير متعامد للمحاور للبيانات الكلية لاستجابات 1174 فرداً عن 89 فقرة ستة عوامل فسر العامل الأول منها 24.34% من التباين، وفسرت جميع العوامل ما قيمته 44.65% من التباين الكلي، وهذا مؤشر على تحقق فرض أحادية البعد، وفق المؤشرات المعتمدة على المكونات الأساسية (Principal Components)، حيث أشار عدد من الباحثين من بينهم ريكاس (Recase, cited in Hattie, 1985) إلى أنه إذا فسر العامل الأول أكبر تباين، فإن هذا مؤشرا على أحادية البعد، وحدد ريكاس تلك القيمة بأن تكون على الأقل 20% من التباين.

كما أظهرت نتائج التحليل أن قيمة الجذر الكامن للعامل الأول 6.81 و للعامل الثاني 1.42 وللعامل الثالث 1.17، في حين كانت قيمة الجذر الكامن للعامل السادس 1.01. وبالاعتماد على مؤشر لورد (Lord, 1980) الخاص بأحادية البعد والذي ينص على أن

الفقرات تكون أُحادية البعد، إذا كانت قيمة الجذر الكامن الأول كبيرة بالمقارنة مع قيمة الجذر الكامن الثاني، وتكون قيمة الجذر الكامن الثاني قريبة من قيمة بقية الجذور الكامنة التي تليه، فإن قيم الجذور الكامنة للعوامل الستة تحقق محك لورد إذ أن قيمة الجذر الكامن الأول 6.81 والثاني 1.42 وبقية الجذور قريبة من قيمة الجذر الكامن الثاني.

وبالاعتماد على محك ديفجي (Divigi, cited in Hattie, 1985) الذي ينص على أن فقرات الاختبار تكون أُحادية البعد، إذ كانت نسبة الفرق بين قيمتي الجذرين الكامنين الأول والثاني إلى الفرق بين قيمتي الجذرين الكامنين الثاني والثالث كبيرة، وإذا نظرنا إلى هذه القيمة نجدها تساوي 21.6 وهي قيمة كبيرة تعبّر بطريقة أخرى عن محك لورد، إذ تبيّن أن قيمة البسط الناتجة من طرح الجذر الكامن الثاني من الجذر الكامن الأول كبيرة نسبياً بالمقارنة مع قيمة المقام الناتجة من طرح قيمة الجذر الكامن الثالث من الجذر الكامن الثاني، حيث تبيّن من مقارنة البسط والمقام والقيمة الناتجة من القسمة أن القيمة الكبيرة للبسط تعود للفرق الكبير بين قيمتي الجذرين الكامنين الأول والثاني وهذا ما يؤكد أن قيمة الجذر الكامن الأول أكبر نسبياً من قيمة الجذر الكامن الثاني وأن القيمة الصغيرة للمقام تعود للفرق القليل بين قيمتي الجذرين الكامنين الثاني والثالث وبالتالي اقتراب القيمتين من بعضهما.

وأكد استخدام الفحص البياني (Scree Plot) لعدد العوامل الناتجة مع قيمة الجذر الكامن للعوامل فرض أُحادية البعد لفقرات الاختبار. بذلك تكون الباحثة قد تأكدت من تحقق افتراض أُحادية البعد لاستجابات 1174 فرداً عن 89 فقرة والعائدة لبيانات الاختبار الكلي.

كما أكدت المحكات التي استخدمتها الباحثة والتي تعتمد على معامل الثبات أيضاً تحقق فرض أُحادية البعد، حيث أظهرت معاملات الارتباط بين الأداء على الفقرة والاختبار بعد حذف الفقرة (Corrected item Total correlation) قيماً مرتفعة اعتبرت مؤشراً على

تحقق أُحادية البعد لبيانات الاختبار الكلي (89 فقرة). وذلك وفقاً لمحك ننلي (Nunnally, cited in Hattie, 1985). كما أظهرت القيم العالية لمعاملات الاتساق الداخلي (كودر – ريتشاردسون (KR-20)) للاختبار الكلي مؤشراً لتحقق فرض أُحادية البعـد وفقاً لرأي كرونباخ (Cronbach, cited in Hattie, 1985). إذ يرى أن معامل الثبات (KR-20) يعد مؤشراً جيداً للتحقق من أُحادية البعد، لأنه يمثل متوسط كل المعاملات النصفية الممكنة (Split-half coefficients) بالإضافة إلى أن (KR-20) تمثل القيمة المتوقعة لنسبة التباين المفسر من العوامل المشتركة بين الفقرات (Common factor) عند ارتباط عينتين عشوائيتين من تجمع الفقرات (Item Pool).

وأكد كرونباخ على أنه حتى يكون الاختبار قابل للتفسير لابد من أن يفسر العامل الرئيس الأول نسبة كبيرة من التباين، وأكد على أن (KR-20) تقدر نسبة التباين والعائدة إلى العوامل المشتركة بين الفقرات.

– مناقشة النتائج المتعلقة بمطابقة الفقرات لنماذج نظرية استجابة الفقرة:

قامت الباحثة بتحليل استجابات 1174 فرداً عن 89 فقرة من نوع الاختيار من متعدد باستخدام برمجية Bilog -Mg لبيانات الاختبار وفق النماذج الثلاثة. وفي ضوء نتائج التحليل تم حذف استجابات الأفراد الذين لم تتطابق استجاباتهم مع الاستجابات المتوقعة تبعاً لقدراتهم، حيث تم حذف استجابات 75 فرداً من أفراد العينة الكلية لم تتطابق استجاباتهم وتوقعات النموذج أُحادي المعلَمة، في حين تم حذف استجابات 74 فرداً لم تتطابق استجاباتهم وتوقعات النموذج ثنائي المعلَمة أما نموذج ثلاثي المعلَمة فقد تم حذف استجابات 98 فرداً. يلاحظ أن عدد الأفراد الذين تم حذف استجاباتهم من النموذج ثلاثي المعلَمة كان الأكبر، في حين تساوى عدد الأفراد الذين حذفت استجاباتهم تقريباً في النموذجين أُحادي وثنائي المعلَمة.

وتفسر الباحثة عدم تطابق استجابات الأفراد لتوقعات النماذج بسبب عدم جدية بعض الطلبة وميلهم إلى الاستجابة عن الفقرات بشكل عشوائي، مما جعل استجاباتهم تختلف عن توقعات النموذج، ويدعم هذا التفسير علامات هؤلاء الطلبة، المنخفضة على الاختبار وفق النموذج ثلاثي المعلَمة، كما أنه من الممكن أن يكون السبب هو ميل بعض الأفراد من ذوي القدرات المتدنية إلى التخمين على الأسئلة التي قيم معلَم الصعوبة لها تفوق قدراتهم. خاصة أن مجموعة كبيرة من الطلبة في أثناء تطبيق الباحثة للاختبار تحدثوا معها بعد انتهاء فترة الاختبار حول صعوبة الاختبار وغرابته بعض الشيء.

وبتفحص قيم معلَم التخمين لفقرات الاختبار وفق النموذج ثلاثي المعلَمة، تتوقع الباحثة أن خروج أفراد من البرنامج نتيجة عدم مطابقة استجاباتهم للنموذج كان بسبب معلَم التخمين، الذي يتيح للأفراد متدني القدرة الوصول للجواب الصحيح عن طريق التخمين، ويؤكد توقع الباحثة هذا أن نسبة كبيرة من الأفراد الذين تم حذف استجاباتهم لعدم تطابقها مع توقعات النموذج ثلاثي المعلَمة كان من ذوي القدرات المتدنية (θ=-4) وإن الخطأ المعياري للتقدير كان عالياً جداً، وعلاماتهم الخام كانت متدنية تراوحت بين 7 و 20 مع العلم بأن العلامة الكلية للاختبار 89.

أما بالنسبة للطلبة ذوي العلامات العالية على الاختبار فتفسر الباحثة سبب حذف البرنامج لاستجاباتهم ربما بسبب سهولة بعض الفقرات قياساً مع الفقرات الأخرى في الاختبار الأمر الذي جعلهم يفكرون بشكل أبعد مما تعنيه الفقرة، مما أدى بهم إلى الإخفاق في التعامل مع الفقرات التي صعوبتها دون مستوى قدراتهم.

وقد يكون سبب آخر خاص بخصائص الأفراد الذين تم حذف استجاباتهم، حيث إن العدد الأكبر من الأفراد الذين تم حذف استجاباتهم كان من الطلبة الذكور في المرحلة الثانوية، حيث تتسم هذه الفئة بشكل عام بالاستهتار حتى في الامتحانات المدرسية، إذ ينظر إليها كثير من الطلبة كفترة استراحة لما قبل الثانوية العامة.

وأما بالنسبة لعدد الفقرات التي لم تطابق الاختبار وفق النموذج أُحادي المعلَمة فقد كان 39 فقرة، في حين كان عدد الفقرات غير المطابقة للاختبار وفق النموذجين ثنائي وثلاثي المعلَمة 19 فقرة و7 فقرات على الترتيب.

يلاحظ هنا وجود علاقة وتيرية متناقصة بين عدد الفقرات المحذوفة وعدد معالم النموذج، وتُرجع الباحثة سبب وجود هذه العلاقة الوتيرية إلى القيود الكثيرة التي يفرضها نموذج أُحادي المعلَمة على بياناته، مقابل التحرر من بعض تلك القيود في النموذج ثنائي المعلَمة، والتحرر الأكبر في النموذج ثلاثي المعلَمة ويشير كل من كيلكار ووايت مان وليشت (Kelkar, Wightman & Luecht, 2000) إلى أن المعلَمين الإضافيين في نموذج ثلاثي المعلَمة يتيحان له فرصة أعلى لمطابقة البيانات الخاصة بالاستجابات لفقرات اختبار من نوع اختيار من متعدد من النماذج الأخرى. وأكد وجهة النظر هذه كلّ من شيرينشنكو وستارك وشان ودراسجو وويليامز وفان ولورد (Chernyshenko, Stark, Chan, Drasgow & Williams, 2001; Fan, 1998; Lord, 1980). وتتفق هذه النتائج مع ما توصلت إليه (جمحاوي، 2000) حين أشارت في نتائج دراستها إلى أن عدد الفقرات المنسجمة مع النموذج يزداد بزيادة عدد معالمه.

وعند دراسة معالِم الفقرات التي حذفت من فقرات الاختبار وفق النموذج ثلاثي المعلَمة نجد أن معلَم التخمين لهذه الفقرات كان أعلى من القيم المقبولة في القياس التربوي حيث تراوحت هذه القيم بين0.307 و 0.769، علماً بأن المتوسط الحسابي لمعلَم التخمين للفقرات المتبقية هو 0.25، وهو يمثل الحد الأعلى لقيمة معلَم التخمين المقبولة لفقرات اختبار من نوع الاختيار من متعدد بأربعة بدائل لكل فقرة منها بديل واحد فقط صحيح (Baker, 2001). ويعزز تلك القيم أيضاً قيم معلَم الصعوبة للفقرات المحذوفة لهذا النموذج حيث تجاوزت قيمة معلَم الصعوبة لثلاث فقرات الواحد صحيح.

وامتازت بعض الفقرات المحذوفة أيضاً بقيم عالية نوعاً ما لمعالم التمييز الخاص بها حيث كانت قيمة معلَم التمييز لفقرة من الفقرات المحذوفة 1.752 ويصف بيكر (Baker, 2001) أن القيم التي تتجاوز 1.7 تعد عالية جداً، أيضاً كانت هناك بعض الفقرات المحذوفة يعد معلَم التمييز لها ضعيف وفق تصنيف بيكر مما جعل تطابقها مع النموذج أيضاً ضعيف.

أما بالنسبة للاختبار وفق النموذج ثنائي المعلَمة فقد امتازت الفقرات المحذوفة بأنها ذات معلَم تمييز يتراوح بين الانخفاض والاعتدال حيث تراوحت قيمه بين 0.307 و 1.133 إذ يشير بيكر إلى أن الفقرات التي تمييزها يتراوح بين 0.35 و 0.64 تعد فقرات ضعيفة التمييز، في حين تعد القيم التي تتراوح بين 0.65 و 1.34 قيماً معتدلة، أي أن النموذج ثنائي المعلَمة قام بحذف بعض الفقرات ذات التمييز الضعيف والمعتدل.

أما ما يتعلق بمعلَم الصعوبة بشكل عام فقد حذفت النماذج الثلاثة فقرات تميّزت بالصعوبة وأخرى تميّزت بالسهولة ويمكن أن يرجع ذلك إلى أن الفقرات الصعبة تمت الإجابة عنها بعشوائية مستخدماً المستجيب التخمين في الوصول للاستجابة الصحيحة، أما فيما يتعلق

بالفقرات السهلة فقد يكون المفحوص ذو القدرة العالية قد فكر بأبعد مما تعنيه الفقرة ويحدث هذا مع المتفوقين ذوي القدرات العالية جداً الذين يميلون إلى تفسير أكثر الفقرات سهولة في الاختبار وتقييمها بأكثر مما تحتمل (Meijer, 1996).

- مناقشة النتائج المتعلقة بصدق الاختبار:

تحققت الباحثة من أن الاختبار يقيس مهارات التفكير الناقد التي صمم لقياسها من خلال التحليل النظري لمفهوم التفكير الناقد والمهارات الأساسية التي تندرج تحته، حيث تم بناء الاختبار اعتماداً على نتائج هذا التحليل. وتحققت الباحثة من الفرضيات التي تبنتها للتحقق من صدق الاختبار، إذ حققت الفرضيتين الأولى والثانية من خلال النتائج التي أظهرها تحليل التباين للأداء على الاختبار الكلي (89 فقرة) لدراسة أثر متغيري الجنس والمستوى الأكاديمي إذ تبين أن هناك فروقاً دالة إحصائياً (α=0.01) لمتغير المستوى الأكاديمي، وأن متوسط الأداء على الاختبار ازداد بزيادة المستوى الأكاديمي وبالتالي بزيادة العمر أيضاً، وتفسر الباحثة ذلك بالاعتماد على أن التفكير الناقد مهارة، وأن هذه المهارة قابلة للتحسين والصقل بزيادة خبرات الفرد، وأن المرء كلما تقدم به العمر والمستوى الأكاديمي تزداد خبراته ويتعرض لمواقف حياتية تضعه أمام خيارات وتحديات تفرض عليه أن يدرس جميع السبل المتاحة أمامه، ويختار الأنسب، مما يساعد على صقل قدراته على الاختيار الأفضل وصنع القرارات والاختيار بين البدائل المتنافسة المتاحة أمامه مما يزيد من قدراته على التفكير الناقد.

تتفق هذه النتيجة مع ما توصل إليه كل من (خليفة، 1990؛ السيد، 1995؛ Frisby, 1991)، حيث أشارت نتائج هذه الدراسات إلى أن متوسطات الأداء لأفراد الدراسة أخذت منحى تصاعدياً في التفكير الناقد بزيادة المستوى الأكاديمي والعمر.

لم تظهر نتائج التحليل أثراً ذا دلالة لمتغير الجنس على الأداء على الاختبارين الكلي (89 فقرة) وهذا ينسجم مع نتائج كثير من الدراسات التربوية التي أجريت محلياً أو عالمياً (الدردور، 2001؛ الشبول، 2004؛ العبد اللات، 2003؛ Burbach, Matkin & Fritz, 2004).

وهذه النتيجة تنسجم مع معطيات مفهوم التفكير الناقد بشكل عام حيث يتعرض كلا الجنسين في المجتمع لنفس المناهج وأنظمة التعليم، وتقريباً نفس ظروف التنشئة الأسرية، مما يجعل قدراتهم تقريباً تنمو بنفس الاتجاه.

ويشير فاسيون وفاسيون (Facione & Facione, 1998) إلى أنه لا يوجد سبب للاعتقاد بأنه هناك اختلاف بين أداء الرجال والنساء على اختبار التفكير الناقد، وأثبتا وجهة النظر هذه من خلال تحليل علامات الطلبة على اختبار كاليفورنيا حيث أظهرت النتائج عدم وجود فروق بين الجنسين.

أما ما يتعلق بنتائج مقارنة المتوسط الحسابي لعلامات طلبة الصف العاشر في مدارس الملك عبد الله الثاني على الاختبار الكلي (89 فقرة) والمتوسط الحسابي لعلامات عينة من طلبة الصف العاشر من ذوي التحصيل المتدني، فقد أظهرت نتائج التحليل فروقاً دالة إحصائياً (α=0.01) ولصالح مدارس الملك عبد الله الثاني للتميّز، وهذه النتيجة توقعتها الباحثة من خلال الفرضية الثالثة التي تبنتها، وذلك لارتباط تعريف التفكير الناقد في هذه الدراسة بالنتاجات المتوقعة من الأشخاص الذين يملكون سمة التميّز، بالإضافة إلى التفوق في التحصيل الدراسي.

كما أظهرت النتائج المتعلقة بمقارنة المتوسط الحسابي لعلامات طلبة كلية الطب على الاختبار الكلي (89 فقرة)، والمتوسط الحسابي لعلامات عينة من طلبة جامعة اليرموك أُخذت من عينة الدراسة، فروقاً دالة إحصائياً (α=0.01) بين المتوسطين ولصالح طلبة كلية الطب.

ويشير جروان (2002ب) بأن الموهوبين والمتفوقين غالباً ما يميلون إلى إظهار تميّزهم في جميع النواحي بما في ذلك التحصيل المدرسي، ويكون تفوقهم واضحاً خاصة في المواد التي تتطلب التفكير المجرد.

تتفق هذه النتيجة مع ما توصل إليه حمادنة (1995)، حيث أشارت نتائج دراسته إلى تفوق أداء الطلبة ذوي التحصيل المرتفع على أداء الطلبة ذوي التحصيل المتدني في كل من اختبار التفكير الناقد الكلي والاختبارات الفرعية. وبلغ معامل الارتباط بين علامات الطلبة على الاختبار بصورته الكلية (89 فقرة)، واختبار كاليفورنيا (2000) المعدل للبيئة الأردنية 0.723، ويعبّر هذا المعامل نوعاً ما عن علاقة معقولة بالمقارنة مع معامل الارتباط بين اختباري كاليفورنيا وواتسون – جليسر للتفكير الناقد حيث بلغت قيمته لمجموعتين 0.405 و 0.544. (Facoine & Facione, 1998) وحيث أن صدق المحك يشير إلى قدرة الاختبار على التنبؤ بمحك خارجي، فإذا كان الاعتقاد بأن التفكير الناقد قادر على التنبؤ بالتحصيل، فإن الباحثة تأمل بأن تكون العلامات على الاختبار ترتبط إيجابياً مع النجاح المدرسي. وبالنظر إلى معامل الارتباط بين علامات نفس طلبة الشعبتين على الاختبار بصورته الكلية ومعدل علاماتهم المدرسية، أشار هذا المعامل إلى علاقة قوية إيجابية (ر=0.834)، وهذا مؤشر قوي لصدق المحك.

– مناقشة النتائج المتعلقة بمؤشرات الثبات:

تراوحت قيم معامل الثبات التي استخرجت للفقرات المشتركة للاختبار (49 فقرة)، وفق النماذج اللوجستية الثلاثة وبأنواع معاملات الثبات الثلاثة التي استخدمت بالدراسة (النظري، الامبيريقي، KR-20) بين 0.856 و 0.908. وتشير هذه القيم نوعاً ما إلى أن الاختبار يتمتع بدقة مقبولة في قياسه لمهارة التفكير الناقد. وتعد هذه القيم عالية إذا ما قورنت

مع قيمة معامل ثبات اختبار كاليفورنيا حيث أظهر الدليل (Manual) لاختبار كاليفورنيا أن قيم KR-20 تراوحت بين 0.68 إلى 0.70. وتتفق هذه النتائج مع نتائج دراسة واجنر وهارفي (Wagner & Harvey, 2003) حيث أشارت نتائج دراستهما إلى أن معامل الاتساق الداخلي لاختبار واتسون – جليسر بصورته الأصلية (α = 0.82) كان أقل من معامل الاتساق الداخلي لاختبار واتسون – جليسر المطوّر وفقاً لنظرية استجابة الفقرة (α = 0.93، وعدد الفقرات 80 فقرة) واختبار مصغر منه (α = 0.90، وعدد الفقرات 40 فقرة). ويشير كل من نوريس وإنيس (Norris & Ennis, cited in Facione & Facione, 1998) إلى أن تقديرات الثبات المتعلقة بالأدوات التي تعمل على اختبار أوجه مهارات التفكير وتتراوح بين 0.65 و 0.75 تعد كافية.

مناقشة النتائج المتعلقة بتقديرات معالم الفقرة:

تراوحت قيم معلم الصعوبة لفقرات الاختبار وفق النموذج أُحادي المعلمة بين –0.985 و 1.658، بمتوسط حسابي 0.125 وانحراف معياري 0.694، في حين تراوحت قيم معلم الصعوبة لفقرات الاختبار وفق النموذج ثنائي المعلمة بين –1.104 و2.050، بمتوسط حسابي 0.121، وانحراف معياري 0.725، وتراوحت قيم معلم الصعوبة لفقرات الاختبار وفق النموذج ثلاثي المعلمة بين –0.162 و 1.798، بمتوسط حسابي 0.799 وانحراف معياري 0.512.

يتبيّن من قيم المدى لمعلم الصعوبة لفقرات الاختبار وفق النماذج الثلاثة أن النموذج ثنائي المعلمة يتمتع بأكبر مدى، حيث غطى مدى واسعاً من متصل السمة مقارنة مع مدى كل من النموذجين أُحادي وثلاثي المعلمة، مما يشير إلى أنه مقياس يقترب من مقياس متساوي الدقة (equal-precise measure) حيث يتسم هذا المقياس حسب ما ورد عن إمبرتسون

ورايس (Embretson & Reise, 2000) بأن منحنى توزيع العلامات الخاص به منبسط نسبياً (Flat) على مدى متصل السمة، وهذا يعني مجموعة من الفقرات ذات التمييز العالي بمدى واسع من معلم الصعوبة.

ولدى مقارنة مدى كل من معلم الصعوبة والقدرة لكل نموذج، يتبيّن أن النموذج ثنائي المعلمة له أكبر مدى لكل من معلمي الصعوبة والقدرة حيث تراوحت قيم معلم الصعوبة بين -1.104 و 2.050 بينما تراوحت قيم معلم الصعوبة للنموذجين أُحادي وثلاثي المعلمة بين -0.985 و 1.658 وبين -0.162 و 1.798 على الترتيب.

وتراوحت قيم معلم القدرة للنموذجين أُحادي وثلاثي المعلمة بين -2.884 و 3.890 وبين -4 و 2.507 على الترتيب في حين تراوحت قيم القدرة للنموذج ثنائي المعلمة بين -2.976 و 3.948.

ولدى الاطلاع على المتوسط الحسابي والانحراف المعياري لمعلم القدرة للنماذج الثلاثة الأُحادي والثنائي والثلاثي حيث كانت (0.364 و 1.696)، (0.359 و 1.716) (-0.063 و 1.973) على الترتيب، وتشير هذه القيم إلى الاقتراب الأكبر للمتوسط الحسابي من القيمة الصفر لمعلم القدرة الخاص بالنموذج ثلاثي المعلمة، مع عدم اقتراب الانحراف المعياري من القيمة واحد، وإذا ما ربطنا هذه القيم مع القيم الخاصة بمعالم الفقرة على النماذج الثلاثة حيث تراوحت قيمة الصعوبة بين -0.985 و 1.658 للنموذج أُحادي المعلمة وبين -1.104 و 2.050 للنموذج ثنائي المعلمة وبين -0.162 و 1.798 للنموذج ثلاثي المعلمة في حين تراوحت قيم التمييز للنموذجين ثنائي وثلاثي المعلمة بين 0.297 و 0.672 وبين 0.412 و 1.489 على الترتيب وهي قيم نوعاً ما قريبة من القيم التي ذكرها كل من هامبلتون وسوامينيثان (Hambleton & Swaminathan, 1985) حيث أشارا إلى أنه إذا كان المتوسط الحسابي للقدرة صفر والانحراف المعياري واحد، فإن قيمة الصعوبة تنحصر بين -2 و 2 والتمييز بين صفر و 2.

وعند تتبع الأخطاء المعيارية في التقدير لمعلم الصعوبة يتبيّن أن أعلى خطأ 0.098 للفقرتين 20 و 55 وصعوبتهما 1.353 و 1.658 وتمثلان أصعب فقرتين في فقرات نموذج أُحادي المعلمة، في حين كان أدنى خطأ 0.084 للفقرات 74 ، 58، 49 حيث قيم الصعوبة لهذه الفقرات 0.125، −0.039، −0.187 على الترتيب، والتي تمثل فقرات متوسطة الصعوبة للمقياس.

وفيما يتعلق بالنموذج ثنائي المعلمة كان أعلى خطأ 0.254 لأصعب فقرة في المقياس حيث صعوبتها 2.050 وهي الفقرة 55 في حين كان أدنى خطأ 0.062 للفقرة التي صعوبتها −0.389 وهي الفقرة 43 وتمثل فقرة متوسطة الصعوبة للمقياس، بشكل عام ارتبط أعلى خطأ مع أصعب الفقرات وأدنى خطأ مع فقرات متوسطة الصعوبة في النموذجين أُحادي وثنائي المعلمة.

أفرز النموذج ثلاثي المعلمة أكبر خطأ معياري 0.398 للفقرة 57 التي صعوبتها −0.023 وهي تمثل فقرة متوسطة الصعوبة، وأدنى خطأ 0.075 للفقرة 69 التي صعوبتها 1.046 وهي تمثل فقرة صعبة نوعاً ما، وأبدت قيم الخطأ المعياري انخفاضاً عند القيم العليا، حيث بلغ 0.123 عند أصعب فقرة ومعامل صعوبتها 1.798، أي يمكن القول أن الخطأ الأعلى ارتبط مع الفقرات متوسطة الصعوبة فيما ارتبط الخطأ الأدنى مع الفقرات الصعبة.

وإذا ما ربطنا بين دقة تقدير معلم الصعوبة التي تم الحصول عليها للنماذج الثلاثة مع طول الاختبار وعدد المفحوصين، يتبيّن أن هذا المعلم يتمتع بدقة مقبولة وذلك وفقاً لما ورد عن سوامينيثان وجيفورد (Swaminathan & Gifford, cited in Hulin, Drasgow & Parsons, 1983) حيث أشارت نتائج دراستهما إلى وجود ارتباطات عالية بين قيم معالم الصعوبة الحقيقية والتقديرية لتوزيعات ملتوية القدرة بلغت 0.8 في حالة اختبار مكون من عشرة فقرات وعدد

المفحوصين 50، ووصلت هذه القيمة إلى 0.95 في حالة عينة كبيرة من المفحوصين عددهم 1000 واختبار طوله 80 فقرة. وعلّق هولين ودراسجو وبارسونز (Hulin, Drasgow & Parsons, 1983) على نتائج هذه الدراسة حين أشاروا إلى أن أثر حجم العينة وطول الاختبار على تقدير معلم الصعوبة كان أقل خطورة مقارنة مع تقديرات معلمي التمييز والتخمين. وعند مقارنة نتائج دراسة سوامينيثان وجيفورد مع نتائج الدراسة الحالية، تبيّن أن توزيعات القدرة ملتوية للنماذج الثلاثة، وأن عدد المفحوصين 1086 مفحوصاً وعدد الفقرات 49 فقرة، مما يشير إلى أن القيم التي أفرزتها بيانات الدراسة الحالية لمعلَم الصعوبة لجميع النماذج تتمتع بدقة مقبولة.

وفيما يتعلق بمعلم التمييز تراوحت قيمه بين 0.297 و 0.672 بمتوسط حسابي 0.449 في النموذج ثنائي المعلمة، في حين تراوحت قيمه بين 0.412 و 1.489 بمتوسط حسابي 0.876 في النموذج ثلاثي المعلمة، واقتربت قيم معلم التمييز للنموذج ثلاثي المعلمة من القيم النموذجية لمعلم التمييز إذ أشار ري (Ree, cited in Hattie, 1984) إلى أن القيم النموذجية لمعلم التمييز تتراوح بين 0.5 و 2.5 بمتوسط حسابي يساوي واحد صحيح، ويضيف ري أن القيم التي أكبر من القيمة اثنين من النادر الحصول عليها.

وحصل روس (Ross, cited in Hattie, 1984) على قيم لمعلم التمييز تتراوح بين 0.47 و 1.99 بمتوسط حسابي 1.28 لاختبار مكون من 20 فقرة من نوع الاختيار من متعدد طبقه على 1000 مفحوص، وحصل على قيم تتراوح بين 0.3 و 1.97 بمتوسط حسابي 1.06 على اختبار آخر مكون من 20 فقرة من نوع الاختيار من متعدد على 1000 مفحوص. وحصل لورد (Lord, cited in Crocker & Algina, 1986) على معلم تمييز يتراوح بين 0.40 و 1.7 بمتوسط حسابي 1.07 وانحراف معياري 0.4.

وبشكل عام أكد كل من كروكر والجينا وهامبلتون وجونز وروجرز وهاتي (Crocker & Algina, 1986; Hambleton, Jones & Rogers, 1993; Hattie, 1984) على أن استخدام عينة كبيرة من المفحوصين يضمن الحصول على دقة كبيرة في تقديرات معلم التمييز.

ولدراسة العلاقة بين قيم معلم التمييز الحقيقية والتقديرة لتوزيعات ملتوية للقدرة، أظهرت نتائج دراسة أجراها كل من سوامينيثان وجيفورد (Swaminathan & Gifford, cited in Hulin, Drasgow & Parsons, 1983) أن معامل الارتباط بين معالم التمييز الحقيقية والتقديرة بلغ 0.82 لاختبار عدد فقراته 80 فقرة بعد تطبيقه على 1000 مفحوص، وهذا يشير أيضاً إلى دقة مقبولة في تقدير معلم التمييز للنموذجين ثنائي وثلاثي المعلمة في اختبار الدراسة الحالية، إذ إن عدد المفحوصين تجاوز 1000 مفحوص وفقرات الاختبار 49 فقرة وأكد كل من لورد وسوامينيثان وجيفورد (Swaminathan & Gifford, cited in Hulin, Drasgow & Parsons, 1983; Lord, 1980) على أن الحصول على تقديرات دقيقة لمعلم التمييز يلزم توفير اختبار مكون من 50 فقرة و 1000 مفحوص، وبينت دراسة هولين (Hulin, cited in Crocker & Algina, 1986) أن هناك اتساقاً كبيراً بين المعالم الحقيقية والتقديرة كلما زاد حجم العينة، وبذا فإن معلمي الصعوبة والتمييز بالاعتماد على ما ورد يحققان دقة مقبولة.

وفيما يتعلق بمعلم التخمين فقد تراوحت قيمه بين 0.101 و 0.473 بمتوسط حسابي 0.245، ويمثل هذا المتوسط تقريباً الحد الأعلى للقيمة المقبولة في القياس التربوي والتي تساوي واحداً مقسوماً على عدد بدائل الفقرة، إذ إن للفقرة أربعة بدائل فإن قيمة المتوسط الحسابي التي تم الحصول عليها في الاختبار تعد نوعاً ما مقبولة.

إن قيمتي التخمين 0.101 و 0.249 قابلتها قيمتان لمعلم الصعوبة 0.635 و 0.610، في حين اتسمت الفقرة التي لها أكبر قيمة لمعلم التخمين 0.473 بالصعوبة إذ إن قيمة معلم الصعوبة 1.074 وذات تمييز معتدل حيث جاءت قيمة معلم التمييز 1.289.

بلغ عدد الفقرات التي تراوحت قيم معلم التخمين لها بين 0.251 و 0.473 (19) فقرة تشكّل 39% من عدد فقرات الاختبار، اتسمت هذه الفقرات نوعاً ما بأنها متوسطة الصعوبة حيث كان هناك 11 فقرة قيمة معلم الصعوبة لكل منها أقل من واحد، وأيضاً كان هناك 15فقرة قيمة معلم التمييز لكل منها أقل من واحد.

ويشير بيكر (Baker, 2001) إلى أن قيم معلم التخمين ليس له تأثير واضح عندما تكون قيم معلم الصعوبة أقل من صفر، وقيم معلم التمييز أقل من واحد. مما يجعل نوعاً ما القيم التي تم الحصول عليها في الدراسة الحالية مقبولة بعض الشيء حيث قيم معلم التمييز أقل من واحد.

ويمكن القول وبشكل تقريبي إن معالم النموذج ثلاثي المعلمة تتمتع باتساق (Consistent) معقول وفق ما ذهب إليه هامبلتون (Hambleton, 1989) في قوله بأنه عندما تكون قيم معلم الصعوبة تتراوح بين –2 و 2 والتمييز بين 0.4 و 2 والتخمين بين صفر و 0.25 فإن المعالم متسقة. وعند مقارنة هذه القيم مع نتائج الدراسة الحالية للنموذج ثلاثي المعلمة، تبين بأن هناك اتساقاً بين تلك النتائج إذ أن قيم معلم الصعوبة تتراوح بين –0.162 1.798 والتمييز بين 0.412 و 1.489 والتخمين بين 0.101 و 0.473.

وفي دراسة لري (Ree, cited in Hulin, Drasgow & Parsons, 1983) أشارت نتائجها إلى أن معامل الارتباط بين قيم معلم التخمين الحقيقية والتقديرية بلغ 0.233 لاختبار ملتو القدرة، و 0.557 لتوزيع متماثل للقدرة، يتكون هذا الاختبار من 80 فقرة تم تطبيقه على 2000 مفحوص، واستنتج الباحث: أنه على الرغم من إجراء الاختبار الطويل ووجود العينة الكبيرة لم تتوفر تقديرات دقيقة لمعلم التخمين.

وتتفق نتائج الدراسة الحالية لقيم معلم التخمين مع القيم التي حصل عليها واجنر وهارفي (Wagner & Harvey, 2003) حيث حصلا على تقديرات لمعلم التخمين بلغ قيمة المتوسط الحسابي لهذه التقديرات 0.262 لفقرات الاختبار الذي طوراه لقياس مهارات التفكير الناقد وفق نظرية استجابة الفقرة.

وبذا ترى الباحثة أن تقديرات المعالم للفقرات التي حصلت عليها للنماذج الثلاثة مقبولة ضمن محكات القياس التربوي.

مناقشة النتائج المتعلقة بالدقة النسبية:

أظهرت النتائج المتعلقة بالدقة النسبية للاختبار الممثل بالفقرات المشتركة (49 فقرة) أن النموذجين الأُحادي والثنائي أبديا تقارباً في قيم المعلومات وبالتالي اقتربت قيم الدقة النسبية من الواحد صحيح حيث تراوحت قيم الدقة بين 1.015 و 1.055 لقيم القدرة الواقعة بين م4، وم90، على متصل القدرة، مما يشير إلى تفوق قليل للنموذج ثنائي المعلمة على أُحادي المعلمة، ويعود ذلك إلى وجود المعلم الإضافي للنموذج ثنائي المعلمة (التمييز) والذي يضيف دقة أكثر نتيجة المعلومات الأكثر التي تقدمها الفقرة، تزداد بزيادة قيمة معلم التمييز حيث كانت القيمة الأكبر للدقة 1.055. وتتفق هذه النتيجة مع وجهات نظر كل من كلكار ووايت مان ولشت وسلايند ولن (Kelkar, Wightman & Luecht, 2000; Slinde & Linn, 1978).

أما عند مقارنة النموذجين أُحادي وثلاثي المعلمة، تبيّن أن النموذج ثلاثي المعلمة لم يقدم أية معلومات تذكر عند المستويات الدنيا للقدرة، حيث اقتربت قيم الدقة النسبية من الصفر. وتعزو الباحثة سبب تفوق النموذج أُحادي المعلمة على ثلاثي المعلمة في المستويات الدنيا للقدرة إلى عامل التخمين، حيث تزداد قيم الخطأ المعياري في تقدير القدرة والناتجة من وصول أفراد متدني القدرة إلى الإجابة الصحيحة عن فقرات تفوق صعوبتها قدراتهم، وبزيادة الخطأ المعياري تقل قيم دالة المعلومات التي يقدمها الاختبار.

وبدراسة مدى قدرة كل نموذج يتبيّن أن أدنى قيمة لمعلم القدرة للنموذج أُحادي المعلمة كانت –2.884 في حين كانت أعلى قيمة 3.890، وتراوحت قيم معلم القدرة للنموذج ثنائي المعلمة بين –2.976 و 3.948 في حين كانت بين –4 و 2.507 للنموذج ثلاثي المعلمة.

تظهر هذه الأرقام للقيم الدنيا لمعلم القدرة للنماذج الثلاثة أن النموذج ثلاثي المعلمة هو الأكثر تطرفاً نحو القيم الدنيا من القدرات، مما يشير إلى وجود عدد من الطلبة فاقت صعوبات الفقرات قدراتهم مما جعل بعضهم يميل إلى التخمين العشوائي للوصول إلى الإجابة، ويؤكد لورد (Lord, 1980) على أنه عند قيام المفحوصين ذوي القدرة المتدنية بالتخمين عشوائياً على الفقرات الصعبة، فإن هذا يؤدي إلى نتائج عشوائية تضعف التفسير، وأشار كل من ألين وين ودي جروجتر وفان در كامب وإمبرتسون ورايس (Allen & Yen, 1979; de Gruijter & Van der Kamp, 2005; Embretson & Reise, 2000) إلى أن أثر التخمين يعمل على خفض كمية المعلومات التي تزود بها الفقرة مع بقاء الأشياء الأخرى ثابتة، وأشار دي جروجتر وفان در كامب إلى دراسة قاما بها بمقارنة دالة المعلومات لثلاثة فقرات حيث قاما بتغيير معلم التمييز مرة، ومعلم التخمين مرة أخرى، وأشارت نتائج الدراسة إلى أن أفضل دالة معلومات كان في حالة a=2, c=o وأن الفرق بين كميتي دالة المعلومات في حالة c=0، a=2, و a=2 ،c=0.25 كان كبيراً عند مستويات القدرة الدنيا، وتدنى هذا الفرق كثيراً عند مستويات القدرة العالية.

وعند دراسة قيم الدقة النسبية للنموذج ثلاثي المعلمة نسبة إلى النموذج أُحادي المعلمة في حالة القيم المتوسطة والعالية للقدرة، تبيّن أن قيم الدقة النسبية بدأت بالازدياد إلى أن وصلت الذروة عند $م_{100}$ حيث بلغت 4.141 أي كأن الاختبار ثلاثي المعلمة يعمل بما يعادل أنه أطول بما يزيد على أربعة أضعاف طول الاختبار أُحادي المعلمة.

وقد أظهرت النتائج ازدياد قيم الدقة النسبية بشكل ملحوظ في النصف الثاني لمتصل القدرة تقريباً، وبالتالي تفوّق نموذج ثلاثي المعلمة على النموذج أُحادي المعلمة وتفــوّق أيضاً على النموذج ثنائي المعلمة لتقارب قيم المعلومات للنموذجين الأُحادي والثنائي من بعضهما، وتفسر الباحثة هذا الأمر بزيادة تعقيد النموذج ثلاثي المعلمة نتيجة وصف الفقرة بمعلمين إضافيين هما التمييز والتخمين. ولدى دراسة قيم معلم التخمين لفقرات النموذج ثلاثي المعلمة يتبيّن أنها نوعاً ما ليست عالية وضمن المحكات التي أوردها الأدب التربوي، حيث بلغ المتوسط الحسابي لمعلم التخمين 0.245، أيضاً يقل أثر معلم التخمين لدى القدرات العليا وفق ما أشار إليه كل من ألين وين ولورد ونوفيك (Allen & Yen, 1979; Lord & Novick, 1968) وهذا ما أكدت عليه نتائج دراسة دي جروجتر وفان در كامب حيث قارنا دالتي معلومات فقرتين اختلفتا في قيم معلم التخمين، وأظهرت النتائج عدم تأثر القدرات العليا بهذا الاختلاف بين قيم معلم التخمين.

وسبب آخر تعزو فيه الباحثة تفوق النموذج ثلاثي المعلمة على النموذجين أُحادي وثنائي المعلمة في كمية المعلومات هو ارتفاع قيم معلم التمييز لفقرات النموذج ثلاثي المعلمة حيث كانت أدنى قيمة لهذا المعلم 0.412 وأعلى قيمة 1.489 في حين بلغت أدنى قيمة لهذا المعلم 0.297 وأعلى قيمة 0.672 في نموذج ثنائي المعلمة، وتتفق هذه النتائج مع النتائج التي عرضها كل من لورد ونوفيك (Lord & Novick, 1968) حين قارنا كمية المعلومات لمجموعة من الاختبارات مختلفة الأطوال ومختلفة أيضاً في قيم معلم التمييز، واستنتجا أن الاختبارات الآتية أدت إلى نفس كمية المعلومات:

- الاختبار 1 عدد فقراته 30 و a=0.10
- الاختبار 2 عدد فقراته 21 و a=0.14
- الاختبار 3 عدد فقراته 15 و a=0.21
- الاختبار 4 عدد فقراته 9 و a=0.33
- الاختبار 5 عدد فقراته 5 و a=0.6
- الاختبار 6 عدد فقراته 3 و a=1

حيث يشير الرمز a إلى معلم التمييز.

وأشارت نتائج دراسة قام بها كل من هامبلتون وسوامينثان (Hambleton & Swaminathan, 1985) إلى أن دالة المعلومات تزداد كلما زادت قيم معلم التمييز، وأيد وجهة النظر هذه كل من بيكـــر وهولين ودراسجو وبارسونز (Baker, 2001; Hulin, Drasgow & Parsons, 1983)، وأشار كل من هامبلتون وجونز وروجرز (Hambleton, Jones & Rogers, 1993) إلى أن معلم التمييز الأكبر للفقرة يؤدي إلى الإسهام الأكبر لتلك الفقرة في الاختبار نتيجة كمية المعلومات الكبيرة التي تقدمها تلك الفقرة. وهذا يفسر تفوق النموذج ثلاثي المعلمة على النموذج أحادي المعلمة عند القدرات العليا حيث وصلت قيمة الدقة النسبية إلى 4.141، أي أن النموذج ثلاثي المعلمة عند المستويات العليا من القدرة قدّم تقريباً معلومات تفوق المعلومات التي قدّمها نموذج أحادي المعلمة بما يعادل ثلاثة أضعاف إلى أربعة أضعاف طول الاختبار أحادي المعلمة، وبالتالي يتبيّن تفوق النموذج الثلاثي المعلمة على كل من النموذجين الأحادي والثنائي المعلمة باستثناء القيم المتدنية للقدرة. مما يجعل من النموذج الثلاثي المعلمة الأصلح للفئة المتوسطة والموهوبة من الأفراد.

وبشكل عام قدم النموذج ثلاثي المعلمة معلومات أكبر من المعلومات التي قدّمها كلاً من النموذجين الثنائي والأُحادي، حيث بلغت أعلى كمية للمعلومات للاختبار وفق النموذج ثلاثي المعلمة 16.994، في حين كانت 6.943 للثنائي و 6.583 للأُحادي. وأشار كل من إمبرتسون ورايس (Embretson & Reise, 2000) إلى أن كمية المعلومات للاختبار التي تكون حول القيمة عشرة، والخطأ المعياري 0.31 تعد كمية معلومات مقبولة، ثم ربط إمبرتسون ورايس (Embretson & Reise, 2000) بين الخطأ المعياري ومعامل الثبات، حيث أشارا إلى أنه في حالة الخطأ المعياري 0.3 فإنه يتناسب مع معامل الثبات 0.9، وبالرجوع إلى معاملي الثبات اللذين قدمهما النموذج ثلاثي المعلمة فقد كانت 0.878 للنظري (Theoretical Reliability) و0.908 للتجريبي (Empirical Reliability).

كما أظهرت النتائج المتعلقة بالبحث عن النموذج الافضل باستخدام مؤشر إمبرتسون ورايس (Embretson & Reise, 2000) أن الاختبار وفق النموذج ثلاثي المعلمة قدّم دقة أكبر وبالتالي معلومات أفضل من معلومات الاختبار وفق النموذج ثنائي المعلمة، حيث كانت قيمة كاي تربيع المحسوبة أكبر من القيمة الحرجة عند مستوى الدلالة ($\alpha = 0.01$). كذلك أشارت النتائج إلى أن قيمة كاي تربيع المحسوبة أكبر من القيمة الحرجة عند استخدام مؤشر إمبرتسون وريس للمقارنة بين الاختبار وفق النموذجين أُحادي وثنائي المعلمة. مما يشير إلى أن الاختبار وفق النموذج ثنائي المعلمة قدّم دقة أكبر وبالتالي معلومات أفضل من النموذج أُحادي المعلمة. كذلك أشارت النتائج إلى أن الاختبار وفق النموذج ثلاثي المعلمة قدّم دقة أكبر بالمقارنة مع الاختبار وفق النموذج أُحادي المعلمة، إذ كانت قيمة كاي تربيع المحسوبة أكبر من القيمة الحرجة عند مستوى الدلالة ($\alpha = 0.01$).

لكن يبقى النموذج أُحادي المعلمة الأفضل في إحدى الحالات الآتية:

- حجم العينة المتاحة للباحث صغيرة.
- الأفراد من ذوي القدرات المتدنية.
- البرامج الحاسوبية المتوفرة للباحث تعطي تقديرات لبرنامج أحادي المعلمة فقط(Hambleton & Swaminathan, 1985).

مناقشة النتائج المتعلقة بالفاعلية النسبية:

أبدت النتائج المتعلقة بالفاعلية النسبية والخاصة بالاختبار المستل من الاختبار الكلي ويمثّل الفقرات السهلة ملائمة هذا الاختبار لذوي القدرات الدنيا للنماذج الثلاثة، حيث بينت النتائج أن الاختبار الممثل بالفقرات السهلة يقدم معلومات أفضل لدى المفحوصين ذوي القدرات الدنيا، وتقل هذه المعلومات عند القدرات العليا، لكن تفوق النموذج ثلاثي المعلمة على بقية النماذج في إعطائه معلومات أكثر عند المستويات الدنيا للقدرة، وتفسر الباحثة هذا الأمر بالمدى الواسع من القدرات الدنيا في النموذج ثلاثي المعلمة حيث كانت أدنى علامة للقدرة -4، مما يزيد فرصة تناسب مستوى صعوبة الفقرات السهلة مع قدرات الأفراد.

تتفق هذه النتيجة مع دراسة أجراها لورد (Lord, 1980) أشارت نتائجها إلى أنه في حالة حذف نصف فقرات الاختبار التي تمثل الفقرات الصعبة فإن مستوى الفاعلية للاختبار الذي يمثل الفقرات السهلة نسبة إلى الاختبار الكلي يتحسن عند مستويات القدرة المتدنية. وأعطى تقريباً نصف دقة المقياس الأصلي.

كما تتفق نتائج الدراسة الحالية مع نتائج دراسة كل من هامبلتون وسوامينثان (Hambleton & Swaminathan, 1985) حين أوجدا الفاعلية النسبية لفقرات مستلة من الاختبار الكلي وتمثل الفقرات السهلة، حيث كانت قيم الفاعلية عالية للفقرات السهلة عند

مستويات القدرة الدنيا، إذ بلغت قيم الفاعلية للاختبار الذي يمثل الفقرات السهلة نسبة إلى الاختبار الكلي 1.58 للاختبار الذي يمثل الفقرات السهلة نسبة إلى الاختبار الكلي عند مستوى القدرة -3.

وفيما يتعلق بالاختبار المستل والذي يمثل الفقرات الصعبة أبدت قيم الفاعلية النسبية انخفاضاً ملحوظاً في القدرات المتدنية، في حين أظهرت ارتفاعاً ملحوظاً عند المستويات العليا من القدرة لجميع النماذج، لكن تفوق النموذج أحادي المعلمة على بقية النماذج إذ أعطى قيماً للفاعلية تراوحت بين 0.544 إلى 0.671 في الفترة الواقعة بين م70 و م98، في حين كانت هذه القيم تتراوح بين 0.418 و 0.679 في النموذج ثنائي المعلمة وكانت بين 0.468 و 0.565للنموذج ثلاثي المعلمة.

وتفسّر الباحثة ازدياد قيم الفاعلية للاختبار الممثل للفقرات الصعبة عند مستويات القدرة العليا، بالتناسب والتوافق بين مستويات صعوبة الفقرات وقدرات الأفراد وأكد هذا التفسير العلاقة الارتباطية العالية بين القدرة والعلامة الخام للاختبار وفق النماذج الثلاثة (الأحادي، والثنائي، والثلاثي) المعلمة حيث كانت على الترتيب (0.995، 0.992، 0.906). اتفقت هذه النتائج أيضاً مع نتائج دراسة كل من هامبلتون وسوامينثان ولورد (Hambleton & Swaminathan, 1985; Lord, 1980) حيث أشارت نتائج كل من الدراستين إلى أن مستوى الفاعلية للاختبار الذي يمثل الفقرات الصعبة نسبة إلى الاختبار الكلي يتحسن عند مستويات القدرة العالية.

وتتفق نتائج هذه الدراسة وبشكل تقريبي مع نتائج دراسة هامبلتون وسوامينثان (Hambleton & Swaminathan, 1985) حيث أظهرت نتائج الدراسة قيماً لدالة المعلومات متدنية لدى القدرة الدنيا لكنها الأعلى في المنطقة المتوسطة من القدرات.

ويبقى القرار لمستخدم الاختبار في اختيار نوع الفقرات بشكل يتلاءم مع مستوى قدرات المفحوصين لديه. حيث تسهل عليه نتائج الدراسة الحالية اختيار أقصر اختبار بالمواصفات التي يريدها وتتفق ومستوى المفحوصين لديه.

مناقشة النتائج المتعلقة بالتحقق من خصائص (ميزات) النموذج:

تراوحت قيم معاملات ارتباط سبيرمان لمعلم الصعوبة للمجموعات المختلفة مأخوذة مثنى مثنى لفقرات الاختبار وفق النموذج أحادي المعلمة بين 0.84 و 0.96 في حين تراوحت هذه القيم بين 0.82 و0.95 لفقرات الاختبار وفق النموذج ثنائي المعلمة، وكان أدنى معامل ارتباط بين المجموعات لفقرات الاختبار وفق النموذج ثلاثي المعلمة 0.72 وبلغ أعلى معامل 0.94.

يتبين من القيم السابقة أن جميع القيم العليا لمعامل الارتباط كانت متساوية للنماذج الثلاثة، وكانت القيمة الأقل للقيم الدنيا للنموذج ثلاثي المعلمة، إلا أن جميع القيم لجميع النماذج تعبّر عن علاقة إيجابية قوية بين قيم معلم الصعوبة لمجموعتي الفقرات السهلة والصعبة.

وبشكل عام أبدت النتائج تحقق معقول لخاصية اللاتغير لمعلم الصعوبة، وتحقق تام لخاصية اللاتغير لمعلمي التمييز والتخمين إذ أشارت قيم معاملات الارتباطات بين المجموعات المختلفة إلى العلاقة الارتباطية التامة ($r = 1$).

وفيما يتعلق بالنتائج الخاصة بالتحقق من خاصية اللاتغير لمعلم القدرة، أظهرت النتائج أن معامل ارتباط القدرة في النموذج أحادي المعلمة كان 0.68 بين الفقرات السهلة والصعبة في حين كان 0.67 في النموذج ثنائي المعلمة و0.65 للنموذج ثلاثي المعلمة، مما يشير إلى أن أكثر النماذج تقارباً بين قيمتي القدرة هو النموذج أحادي المعلمة.

وتتفـــــق نتائج هذه الدراسة بشكل تقريبي مع نتائج دراسة كيلكار ووايت مان وليشت (Kelkar, Wightman & Luecht, 2000)، حيث أشارت نتائج دراستهم إلى أن معاملات الارتباط للقدرات بين المجموعات المختلفة تراوحت بين 0.6 إلى 0.76 للنموذج أُحادي المعلمة وبين 0.61 إلى 0.77 للنموذج ثنائي المعلمة وبين 0.62 إلى 0.77 للنموذج ثلاثي المعلمة، لكن اختلفت نتائج دراسة كيلكار ووايت مان وليشت مع نتائج الدراسة الحالية في أن النموذج ثلاثي المعلمة في الدراسة الحالية كان الأقل تمتعاً بخاصية اللاتغير لمعلم القدرة. في حين لم تظهر هذه النتيجة في دراسة كيلكار ووايت مان وليشت (Kelkar, Wightman & Luecht, 2000).

بشكل عام أشارت النتائج إلى توفر خاصية اللاتغير بدرجة بسيطة، وربما يعود عدم التطابق التام في تقديرات القدرة ومعلم الصعوبة إلى أمور عدة منها:

- الخطأ العيني.
- عدم الجدية أو عدم بذل جهود عالية في الإجابة كون هذا الاختبار ليس من اختبارات أعلى أداء ولا تمثل علامته أي شيء بالنسبة للطالب.
- عدم مطابقة البيانات للنموذج بشكل تام.
- عدم تحقق افتراضات النموذج بشكل تام إذ إن الباحثة استخدمت مؤشرات تشير إلى إمكانية التحقق ولا تجزم بتحقق الافتراض، واعتبرت الباحثة أن تحقق أُحادية البعد يؤدي إلى تحقق الاستقلال الموضعي وقد يكون في ذلك تجاوز.

مناقشة النتائج المتعلقة بالمعايير والرتب المئينية والمؤشرات الإحصائية الوصفية للاختبار الكلي والاختبار بصورته النهائية:

أظهرت النتائج أن المتوسطات الحسابية للمجموعات المختلفة للاختبار الكلي كانت متسقة مع الأدب التربوي إذ إن المتوسط الحسابي لأداء طلبة المرحلة الأساسية العليا على الاختبار 38.82 والثانوية 46.74 وطلبة الجامعة 55.02، علماً بأن العلامة الكلية للاختبار 89. حيث تشير هذه القيم إلى العلاقة المتزايدة وتيرياً بين المستوى التعليمي للمجموعات المختلفة لعينة الدراسة والمتوسط الحسابي للأداء على الاختبار لهذه المجموعات.

كما بيّنت المؤشرات الإحصائية الوصفية لبيانات الاختبار الكلي والاختبار بصورته النهائية أن الالتواء كان موجباً للبيانات الخام للمرحلتين الأساسية والثانوية في حين كان سالباً للمرحلة الجامعية، مما يشير إلى أن مستوى صعوبة فقرات الاختبار يعد أعلى من قدرات الأفراد للمرحلتين الأساسية العليا والثانوية، وهذا ما لمسته الباحثة في أثناء تطبيقها الاختبار على هاتين الفئتين.

كما بينت نتائج الإحصاء الوصفي لبيانات الاختبار الكلي تقارب قيم مقاييس النزعة المركزية من بعضها لجميع المجموعات، لكن كانت مجموعة طلبة المرحلة الثانوية هي الأكثر تقارباً؛ لأن مقدار المتوسط الحسابي 46.74، والوسيط 45 والمنوال 45.

وأبدت نتائج المؤشرات الإحصائية الوصفية لبيانات الاختبار بصورته النهائية تقارباً بين المتوسط الحسابي والوسيط لكن كان هناك تباعد بين المتوسط والمنوال.

وقد أبدت المجموعات تقارباً في قيم الانحرافات المعيارية حيث كان 12.40 لطلبة المرحلة الأساسية العليا و13.83 لطلبة المرحلة الثانوية في حين كان 14.85 للطلبة الجامعيين لبيانات الاختبار الكلي، وكانت الانحرافات المعيارية لبيانات الاختبار بصورته النهائية 11.45 و 12.92 و 14.04 للمجموعات الثلاثة: طلبة المرحلة الأساسية العليا وطلبة المرحلة الثانوية وطلبة المرحلة الجامعية على الترتيب.

اتسمت قيم المئينيات بالتزايد بالانتقال من المرحلة الأساسية العليا إلى الثانوية إلى المستوى الجامعي ولجميع قيم المئينيات من م$_{1}$ إلى م$_{100}$ للاختبار الكلي والاختبار بصورته النهائية، لكن كان مقدار الزيادة في القيم المتطرفة للمئين أكثر تقارباً حيث كانت قيمة العلامات الخام عند م$_{100}$ هي 80، 82، 84 للمجموعات الثلاث: طلبة المرحلة الأساسية العليا، وطلبة المرحلة الثانوية، وطلبة المرحلة الجامعية على الترتيب، في حين ظهرت العلامات الخام 30، 36، 43 أيضاً للمجموعات الثلاث على الترتيب عند م$_{25}$، وكانت العلامات الخام 37، 46، 54 عند م$_{50}$ لبيانات الاختبار الكلي.

بينما كانت قيم العلامات الخام عند م$_{100}$ 73، 76، 77 للمجموعات الثلاث: طلبة المرحلة الأساسية العليا، وطلبة المرحلة الثانوية، وطلبة المرحلة الجامعية على الترتيب، في حين ظهرت العلامات الخام 28، 33، 42 عند م$_{25}$، وكانت العلامات الخام 34، 42، 52 عند م$_{50}$ لبيانات الاختبار بصورته النهائية.

تتفق هذه النتائج بشكل عام مع نتائج المئينات التي وردت في دليل اختبار واتسون وجليسر (Watson & Glaser,1964) حيث أبدت القيم تزايداً بالانتقال من الصف التاسع لغاية الثاني عشر، كذلك تزايداً بالانتقال من م$_{2}$ إلى م$_{99}$.

كانت نتائج الإناث في المرحلة الأساسية العليا أعلى من نتائج الذكور في نفس المرحلة، وتقاربت القيم في المرحلة الثانوية، وأبدى الذكور تفوقاً على الإناث في المرحلة الجامعية، لكن بشكل عام بينت نتائج التحليل عدم وجود فروق ذات دلالة في الأداء يعود للجنس، وذلك في بيانات الاختبار الكلي.

تتفق هذه النتائج بشكل عام مع النتائج التي أوردها دليل اختبار كاليفورنيا للتفكير الناقد فاسيون و فاسيون (Facione, and Facione, 1998) حيث أشار الدليل إلى أنه لا يوجد فروق ذات دلالة بين أداء الذكور والإناث على الاختبار.

بلغ المتوسط الحسابي للأداء على الاختبار الكلي لمجموعات الدراسة الثلاث بشكل عام 44.593 والانحراف المعياري 14.350، في حين بلغ المتوسط الحسابي لمجموعات الدراسة الثلاث: 38.820 لطلبة المرحلة الأساسية بانحراف معياري 12.400 و46.744 بانحراف معياري 13.823 لطلبة المرحلة الثانوية و55.017 بانحراف معياري 14.851 لطلبة المرحلة الجامعية، في حين بلغ المتوسط الحسابي للأداء على الاختبار بصورته النهائية 41.18 والانحراف المعياري 13.41، في حين بلغ المتوسط الحسابي لمجموعات الدراسة الثلاثة: 35.91 لطلبة المرحلة الأساسية بانحراف معياري 11.45 و 42.99 لطلبة المرحلة الثانوية بانحراف معياري 12.92 و 52.28 بانحراف معياري 14.04 لطلبة المرحلة الجامعية.

ولدى مقارنة هذه القيم مع المتوسطات الحسابية والانحرافات المعيارية لمجموعات مشابهة في دليل واتسون – جليسر النموذج YM (Watson & Glaser, 1964)، كان المتوسط الحسابي لطلبة الصف التاسع 57.7 والعاشر 61.7 والانحراف المعياري لكل صف 11. علماً بأن العلامة القصوى على الاختبار 100. وبلغ المتوسط الحسابي لطلبة الصف

الحادي عشر 64.4 بانحراف معياري 11 وبلغ المتوسط الحسابي لطلبة الكلية 74.4 بانحراف معياري 9.6. وأظهر الدليل قيماً للمتوسطات الحسابية والانحرافات المعيارية خاصة بالمجموعات المعيارية للنموذج ZM حيث كان المتوسط الحسابي لطلبة الصف التاسع 53.6 والانحراف المعياري 9.4، وكان المتوسط الحسابي لطلبة الصف العاشر 56.4 وبانحراف معياري 9.8 ولطلبة الصف الحادي عشر بلغ المتوسط الحسابي 58.8 والانحراف المعياري 10.3 (Watson & Glaser, 1964).

يتبيّن من القيم أعلاه أن قيم النموذج ZM أكثر تقارباً مع المتوسطات الحسابية للمجموعات المعيارية المتناظرة في الدراسة الحالية، وأبدى النموذج YM قيماً أعلى، لكن اتسمت النتائج بالاتساق لجميع المجموعات سواء في اختبار الدراسة الحالية أم في اختبار واتسون– جليسر بنموذجيه.

وبالإطلاع على دليل اختبار كاليفورنيا والمجموعات المعيارية التي أوردها فقد تراوحت المتوسطات الحسابية بين 14.647 و 19.01 علماً بأن العلامة الكلية 34. وبلغ المتوسط الحسابي للأداء بشكل عام في اختبار كاليفورنيا 17.52 للنموذج A في حين بلغ 17.49 للنموذج B وهي قيم مقاربة للقيم الواردة في نتائج الدراسة الحالية، وبلغت أعلى علامة 27 و 28 للنموذجين A, B على الترتيب وأدنى علامة 9 و 7. علماً بأن العلامة الكلية على الاختبار 34 (Facione, & Facione 1998).

وبشكل عام فقد خرجت الدراسة باختبار للتفكير الناقد مبنياً وفق نظرية استجابة الفقرة، مكوناً من 82 فقرة يقيس المهارات الأساسية التي بنت الباحثة الاختبار على أساسها، وقد طابقت هذه الفقرات معايير النموذج ثلاثي المعلمة لنظرية استجابة الفقرة، وتمتع بخصائص سيكومترية مقبولة ضمن المحكات التي أوردها الأدب التربوي في القياس.

التوصيات:

في ضوء النتائج التي تم التوصل إليها في هذه الدراسة يمكن الخروج بالتوصيات الآتية المتعلقة بالتطبيقات التربوية، وبالبحث العلمي في مجال القياس والتقويم.

- نظراً لتمتع الاختبار المبني والمطابق للنموذج اللوجستي الثلاثي المعلمة بمعايير تتفق ومتطلبات النموذج الثلاثي، ونظراً لأنه كان الأفضل فإن الباحثة توصي باستخدام الاختبار بصورته النهائية (82 فقرة) أو أي اختبار فرعي مستل منه لقياس مهارات التفكير الناقد ولأغراض البحث العلمي.
- إجراء المزيد من الدراسات تتعلق ببناء اختبار للتفكير الناقد يتناول أنواعاً أخرى من الفقرات مثل الفقرات المقالية، أو فقرات صواب – خطأ متعدد وغيرها، وفق النظرية الحديثة في القياس، حيث تراعي هذه الأنواع من الفقرات المعرفة الجزئية وتقلل من أثر التخمين.

المراجع

المراجع العربية:

بشارة، موفق سليم .(2003). *أثر برنامج تدريبي لمهارات التفكير عالي الرتبة في تنمية التفكير الناقد والإبداعي لدى طلاب الصف العاشر الأساسي*. رسالة ماجستير غير منشورة، الجامعة الأردنية، عمان، الأردن.

جبريل، منير وجوابرة، سمير. (2003). *كيف تطور مهارات التفكير العليا الإبداعي والناقد لطلبة الصف التاسع في موضوع الهندسة التحليلية*. دائرة التربية والتعليم، مركز التطوير التربوي، القدس.

جروان، فتحي .(2002ب). *أساليب الكشف عن الموهوبين ورعايتهم*، (ط.1). عمان، الأردن: دار الفكر للطباعة والنشر والتوزيع.

جروان، فتحي. (2002أ). *تعليم التفكير، مفاهيم وتطبيقات* (ط.1). عمان، الأردن: دار الفكر للطباعة والنشر والتوزيع.

جمحاوي، إيناس .(2000). *مقارنة خصائص الفقرات وفق النظرية التقليدية ونظرية استجابة الفقرة في مقياس للقدرة الرياضية*. رسالة ماجستير غير منشورة، جامعة اليرموك، اربد، الأردن.

حبيب، مجدي عبد الكريم .(1995). *دراسات في أساليب التفكير*، (ط.1). القاهرة: مكتبة النهضة المصرية.

حمادنة، أحمد فواز .(1995). *مستوى التفكير الناقد في الرياضيات عند طلبة الصف العاشر في الأردن*. رسالة ماجستير غير منشورة، جامعة اليرموك، اربد، الأردن.

الخطيب، مها أحمد حسين .(1993). *أثر كل من درجة الاستقلال المعرفي والتحصيل، والجنس على قدرة التفكير الناقد للفئة العمرية (11-14) سنة في المدارس الحكومية لمنطقة عمان الأولى*. رسالة ماجستير غير منشورة، الجامعة الأردنية، عمان، الأردن.

خليفة، غازي جمال .(1990). *تطوير مناهج الجغرافيا للمرحلة الثانوية في الأردن لتنمية التفكير الناقد والاتجاه نحو المادة*. رسالة دكتوراه غير منشورة، جامعة عين شمس، القاهرة.

الخوالدة، محمد عبد الله .(2002). *أثر توظيف الأحداث الجارية في تنمية مهارات التفكير الناقد لدى طلبة الصف العاشر الأساسي في مبحث التاريخ*. رسالة ماجستير غير منشورة، جامعة اليرموك، اربد، الأردن.

الدردور، عامر محمد .(2001). *أثر استخدام الخرائط المفاهيمية في تنمية التفكير الناقد لدى طلبة الصف السادس الأساسي*. رسالة ماجستير غير منشورة، جامعة اليرموك، اربد، الأردن.

دي بونو، إدوارد. (1997). *التفكير الإبداعي* (ترجمة: خليل الجيوسي ط.1). أبو ظبي، الإمارات العربية المتحدة: منشورات المجمع الثقافي.

ديوي، جون .(1960). *البحث عن اليقين* (ترجمة أحمد فؤاد الأهواني). القاهرة: عيسى البابي الحلبي.

الربضي، مريم سالم. (2004). *أثر برنامج تدريبي قائم على مهارات التفكير الناقد في اكتساب معلمي الدراسات الاجتماعية في المرحلة الثانوية في الأردن لتلك المهارات ودرجة ممارستهم لها*. أطروحة دكتوراه غير منشورة، جامعة عمان العربية، عمان، الأردن.

السيد، عزيزة. (1995). *التفكير الناقد، دراسة في علم النفس المعرفي*. القاهرة: دار المعرفة الجامعية.

الشبول، أريج صالح. (2004). *أثر التعليم باللعب في تنمية مهارات التفكير الناقد في مبحث الجغرافية لدى طلبة الصف العاشر الأساسي*. رسالة ماجستير غير منشورة، جامعة اليرموك، اربد، الأردن.

شطناوي، محمد خالد .(2003). *تقنين اختبار كاليفورنيا لمهارات التفكير الناقد للطلبة الجامعيين في الأردن*. رسالة ماجستير غير منشورة، جامعة مؤتة، الأردن.

الصويتي، رولا عمر. (2001). *أثر استخدام الجزء الأول والجزء الخامس من برنامج كورت لتعليم التفكير في تنمية التفكير الناقد لدى طالبات الصف العاشر الأساسي*. رسالة ماجستير غير منشورة، جامعة اليرموك، اربد، الأردن.

العبد اللات، سعاد إسماعيل .(2003). *أثر برنامج تدريبي مبني على التعلّم بالمشكلات في تنمية مهارات التفكير الناقد لدى طلبة الصف العاشر الأساسي*. أطروحة دكتوراه غير منشورة، جامعة عمان العربية، عمان، الأردن.

عبد الله، محمد قاسم. (شباط 2003). سيكولوجية الذاكرة، *عالم المعرفة*، 290. الكويت: المجلس الوطني للثقافة والفنون.

علام، صلاح الدين .(1986). *تطورات معاصرة في القياس النفسي والتربوي*. الكويت: كلية الآداب.

مارزانو، روبرت وبراندت، روناس وسوهيوز، كارولين وجونز، بوفلاي وبرسيس، برباراز ورانكن، سيتيوات وآخرون .(2004). *أبعاد التفكير، إطار عمل للمنهج وطرق التدريس* (يعقوب حسين نشوان ومحمد صالح خطاب، مترجمون). عمان، الأردن: دار الفرقان. (نشر العمل الأصلي سنة 1988).

وزارة التربية والتعليم. (1994). **ندوة التطوير التربوي**، *رسالة المعلم*، 35 (2).

المراجع الأجنبية:

Allen, Mary J. & Yen, Wendy M. (1979). *Introduction to measurement theory.* California: Brooks/ Cole publishing company.

Alter, Catherine & Egan, Marcia. (1997).Logic modeling: A tool for teaching critical thinking in social work practice. *Journal of Social Work Education, 33* (1), 85-103.

Astleitner, Hermann. (2002). Teaching critical thinking online. *Journal of Instructional Psychology, 29*(2), 53-76.

Baker, Frank B. (2001). *The basics of item response theory.* (2nd ed.). Retrieved October 10, 2002, from: www.ericae.net/irt.

Beyer, Barry. K. (1985). Critical thinking what is it?. *Social Education, 49*(4), 270-276.

Black, Susan. (2004). Habits of thought. *American School Board Journal 191*(12), 52-54.

Brookfield, Stephen D. (1997). Assessing critical thinking. *New Directions for Adult & Continuing Education, 75*, 17-29.

Brown, Harold & Salisch, Sandra. (1996). Clustering. *College Teaching, 44* (1), 29-34.

Browne, M Neil & Keeley, Stuart M. (1988). *Do college students know how to "think critically" when they graduate?.* Southeast Missouri state university, Cape, Girardeau; Research Serving Teaching, 1(9). (ERIC Document Reproduction Service No. ED 298442).

Burbach, Mark E, Matkin, Gina S & Fritz, Susan M. (2004). Teaching critical thinking in an introductory leadership course utilizing active learning strategies: A confirmatory Study. *College Student Journal, 38*(3), 482-493.

Celuch, Kevin & Slama, Mark. (1999). Teaching critical thinking skills for the 21 century: An adversting principles case study. *Journal of Education for Business, 74* (3), 134-140.

Chaffee, J. (1992). Teaching critical thinking across the curriculum. *New Directions for Community Colleges, 20*(1), 25-35.

Chalupa, M. R. (1992). Critical thinking getting minds to work. *Business Education Forum, 47*(1), 21-24.

Chernyshenko, Oleksandr, S, Stark, Stephen, Chan, Kim- Yin, Drasgow, Fritz & Williams, Bruce. (2001). Fitting item response theory models to two personality inventories: Issues and insights. *Multivariate Behavioral Research, 36*(4), 523-562.

Crocker, Linda & Algina, James. (1986). *Introduction to classical & modern test theory*. New York: Holt, Rinehart & Winston.

de Gruijter, Dato N.M. & Van der Kamp, Leo, J. Th. (2005). *Statistical test theory for education and psychology*. Retrieved December 30, 2005 from: www.leidenuniv.n1/~gruijterdnmde.

Dean, Donny Matthew. (2004). An evaluation of the use of web-enhanced home work assignments in high school biology classes (Doctoral dissertation, University of Alabama, 2004). *Dissertation Abstracts International, A65/04*, 130b.

DeManuele, Janine Mary. (2002). Evaluation of the philosophy for children program: A novel approach to critical thinking (Doctoral dissertation, Saint Louis University, 2001). *Dissertation Abstracts International, 62*, 2960.

Dewey, John. (1930). *Human nature and conduct, an introduction to social psychology*. New York: The modern Library.

Elder, Linda & Paul, Richard. (2001). Critical Thinking: Thinking to some purpose. *Journal of Developmental Education, 25*(1), 40-42.

Eldman, Laird Roy Oakes .(2002). The Minnesota test of critical thinking: Structure & psychometric properties. (Doctoral dissertation, University of Minnesota, 2002). *Dissertation Abstracts international, A 63/03*, 859.

Embretson, Susan E, & Reise, Steven P. (2000). *Item response theory for psychologists*. New Jersey: Lawrence Erlbaum associates, inc.

Ennis, Robert. H. (1985). A logical basis for measuring critical thinking skills. *Educational Leadership, 43*(2), 44-48.

Ennis, Robert H. (1993). Critical thinking Assessment. *Theory Into Practice, 32*(3), 179-186.

Facione, P. & Facione, N. (1998). *California Critical thinking skills test form A & form B: test manual.* California Academic Press. USA.

Fan, X. (1998). Item response theory and classical test theory: An empirical comparison of their item / person statistics. *Educational & Psychological Measurement, 58*(3), 375-381.

Frisby, Craig L. (1991). A meta –analytic investigation of the relationship between grade level & mean scores on the Cornell Critical thinking test, level X. *Measurement & Evaluation in Counseling & Development, 23* (4), 162-170.

Gadzella, B.M, Hartsoe, K., & Harper, J. (1989). Critical thinking & mental ability groups. *Psychological Reports, 65,* 1014-1026.

Gadzella, B.M, & Masten, William G. (1998). Critical thinking & learning processes for students in two major fields. *Journal of Instructional psychology, 25*(4), 256-261.

Gadzella, B.M, Stacks, James, Stephens, Rebecca C., & Masten, William G. (2005). Wasten- Glaser critical thinking appraisal, form S, for education majors. *Journal of Instructional Psychology, 32*(1), 9-12.

Gelder, Tim .(2005). Teaching critical thinking. *College Teaching, 53*(1), 41-46.

Gilser, P. (1997). *Digital literacy: The thinking and survival skills new users to make the internet personally & professionally meaning.* New York: Wiely.

Goyne, June Stockes. (2001). Critical thinking in nursing: Perceptions of nursing educators (Doctoral dissertation, AuBurn University, 2001). *Dissertation Abstracts International*, *62*, 2045.

Grow, W. Linda (Ed.). (1989). *Enhancing critical thinking in the Science* (2nd ed.). Huoston: Baylor college of Medicine.

Halford, Graeme S. (1982). *The development of thought*. New Jersey: Lawrence Erlbaum Associates, Inc.

Halpern, Diane, F. (2001). Why wisdom. *Educational Psychologist, 36*(4), 253-256.

Hambleton, Ronald K. (1989). Principles and Selected applications of item response theory, In Linn, Robert, L. (Ed.), *Educational Measurement* (3rd ed., PP. 147-201). New York: American Council on Education, Macmillan Publishing Company.

Hambleton, Ronald K, Jones, Russell W & Rogers H. Jane .(1993). Influence of item parameter estimation errors in test development. *Journal of Educational measurement, 30*(2), 143-155.

Hambleton, Ronald K. & Swaminathan, Hariharan. (1985). *Item response theory: principles and applications*. Boston: Kluwer, Nijhoff Publishing.

Harrigan, Anne & Vincenti, Virgina .(2004). Developing higher –order thinking through an intercultural assignment. *College Teaching 52*(3), 113-120.

Hattie, John .(1984). An empirical study of various indices for determining unidimensionality. *Multivariate Behavioral Research, 19*, 49-78.

Hattie, John. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied psychological measurement, 9*(2), 139-164.

Hulin, Charles L, Drasgow, Fritz & Parsons, Charles K. (1983). *Item response theory, Application to psychological measurement*. Homewood, Illinois: Dow Jones- Irwin.

Irving, J. A & Williams, D. I. (1995). Critical Thinking & Reflective Practice in Counseling. *British Journal of Guidance Counseling, 23* (1), 107-114.

Jacobs, Stanley, S. (1995). Technical characteristics & some correlates of the California Critical Thinking Skills test forms A & B. *Research in higher education, 36*(1), 89-108.

Jacobs, Stanley S. (1999). The equivalence of forms A & B of the California Critical Thinking Skills Test. *Measurement and Evaluation in Counseling & Development, 31*(4), 211-223.

Jenkins, Elizabeth. (1998). The significant role of critical thinking in predicting auditing students' performance. *Journal of Education for Business, 73* (5), 274-280.

Katsioloudes, Marios J. & Tischio, Victoria. (2001). Critical thinking in nonprofit management education. *Human Systems Management, 20*(1), 47-58.

Kelkar, Vinaya, Wightman, Linda. F. Luecht, Richard M. (2000). *Evaluation of the IRT parameter invariance property for the MCAT.* Paper presented at the Annual meeting of the National Council on Measurement in Education as part of the symposium entitled. New Orleans, LA, April, 25-27, 2000, the Graduate school research.

Kerka, Sandra. (1992). *Higher order thinking skills in vocational education.* (ERIC Document Reproduction Service No. ED 350487).

Kreber, Carolin. (1998). The relationships between self-directed learning, critical thinking and Psychological type, and studies in higher education. *Higher Education, 23* (1), 71-87.

Landis, Richard E & Michael, William B. (1981). The factorial validity of three measures of critical thinking within context of Guilford's structure of intellect model for sample of ninth Grad students. *Educational & Psychological Measurement, 41*(4), 1147-1166.

Lipman, Mathew .(1988). *Misconceptions in teaching for critical thinking,* Resource publication, series, No. 3: Montclair state coll. Upper Montclair, NJ. Ins. For critical thinking (ERIC Document Reproduction Service No. ED 352335).

Loo, Robert & Thrope, Kavran. (1999). A sychosomatic investigation scores on the Watson – Glaser critical thinking appraisal, New form S. *Educational & Psychological Measurement, 59*(6), 995-1003.

Lord, Frederic M. (1980). *Applications of item response theory to practical testing problems*. New Jersey: Lawrence Erlbaum Associates, publishers Hillsdale.

Lord, Frederic M. & Novick, Melvin R. (1968). *Statistical theories of mental test scores*. New York: Addison- Wesley Publishing company, Inc.

McCarthy, Mary – Claire .(2004). Critical thinking disposition, belief perspective, and academic performance: Examining relationships (Doctoral dissertation, University of AuBurn, 2004). *Dissertation Abstracts International, A 65/02*, 399.

McGuinness, Carol. (1993). Teaching thinking: New signs for theories of cognition. *Educational Psychology, 13* (3/ 4) 305-309.

McPeck, John E. (1990). Critical thinking and subject specificity: A Reply to Ennis. *Educational Researcher, 19*(4), 10-12.

McWhorter, Kathleen T. & Collins Harper. (1992). *Study & thinking skills in college* (2nd ed.). Authors.

Meijer, Rob R. (1996). Person – Fit research: An introduction. *Applied Measurement in Education, 9*(1), 3-8.

Modjeski, Richard. B. & Michael, William B. (1983). An evaluation by panel of psychologist of the reliability & validity of two tests of critical thinking. *Educational & Psychological Measurement, 43*, 1187-1197.

Onwuegbuzie, Anthony J. (2001). Critical thinking skills: A comparison of Doctoral and Master's – Level student's. *College Student Journal, 35*(3), 477-480.

Ormrod, Jeanne Ellis. (1995). *Educational psychology: principles & applications* (1st ed.). New Jersey: Prentice-Hall, Inc.

Paul, Richard. (1998). Critical thinking: placing it at the heart of ethics, instruction. *Journal of Development Education, 22* (2), 36-38.

Paul, Richard, Elder Linda & Bartell, Ted. (2004). A Brief of history of the ideal critical thinking. *California teacher preparation for instruction*. Retrieved April 15, 2005 from: www.criticalthinking.org.

Paul, Richard & Nosich, Gerlad. (1992). *A model for the national assessment of higher order thinking*. ERIC Document Reproduction Service No. ED 353296).

Petress, Ken .(2004). Critical thinking: An extended definition. *Education, 124*(3), 461-466.

Pike, Gary R. (1997). The California critical thinking skills test. *Assessment Update, 9*(2), 10-11.

Polette, N. (1982). *3R's for the gifted: Reading, writing, & research*. Littleton, Co: Libraries Un Limited.

Potts, Bonnei. (1994). *Strategies for teaching critical thinking*. (ERIC Document Reproduction Service No. ED 385606).

Royalty, Joel. (1995). The generalizability of critical thinking: Paranormal beliefs versus statistical reasoning. *Journal of Genetic Psychology, 156*(4), 477-489.

Salder, Willbert L. (1993). Awakening students critical thinking power through logic problems. *Journal of Instruction Psychology, 20*(4), 354-367.

Sawin, Gregory. (2004). General semantics as critical thinking: A personal view, Etc. *Review of General Semantics, 61*(2), 238-292.

Schuck, Robert William. (2003). Responses to the use of the internet with California standards – based middle school algebra curriculum among at risk, Latino students (Doctoral dissertation, University of southern California, 2004). *Dissertation Abstracts International, A 64/12*, 4345.

Scriven, L & Paul R. (1992). *Defining critical thinking. National Council for Excellence in Critical Thinking instruction*. Retrieved, March 30, 2005, from: www.critcalthinking.org.

Siegel, Harvey. (1999). What (Good) are thinking dispositions. *Educational theory*, *49* (2), 207-223.

Slinde, Jefferey A. & Linn, Robert L. (1978). An exploration of the adequacy of the Rasch model for the problem of vertical equating. *Journal of Educational Measurement, 15*(1), 23-35.

Sormunen, Carolee & Chalupa, Marilyn. (1994). Critical thinking skills research: Developing evaluation techniques. *Journal of Education for Business, 9*(3), 172-178.

Spicer, Karin – Leigh & Hanks, William E. (1995). *Multiple measures of critical thinking*. skills and predispositions in assessment of critical thinking (ERIC Document Reproduction Service No. ED 391185).

Stein, Barrys; Haynes, Adaf. & Unterstein, Jenny. (2003). *Assessing critical thinking. Paper accepted for presentation at SACS annual meeting in Nashville, Tennessee in December 2003*. http://web.tntech. edu/sti/SACS% 20 presentation%paper.pdf.

Sternberg, Robert J. (2004). Four alternative futures for education in the United States: It's our choice. *School Psychology Review, 33*(1). 67-77.

Temple, Charles. (2000). What can we learn from (15.000) teachers in Central Europe & Central Asia. *Reading teacher, 54*(3), 312-315.

The Role of Critical Thinking (1997). *Ardel Wellness Report, 47*, 1-3.

Udall, Anne, J & Daniels, Jones E. (1991). *Creating the thoughtful classroom :Strategies to promote student thinking, Crades 3-12.* Tucson: Zephyer press.

Wagner, Teresa A. & Harvey, Robert J. (2003). *Developing A new critical Thinking Test Using item Response Theory. Paper presented at the 2003 Annual conference of the Society for Industrial Organizational Psychology.* Orlando. Retrieved May, 4, 2005, from: www.criticalthinking.org.

Walkner, Paul & Finney, Nicholas. (1999). Skill development & critical Thinking in higher education. *Teaching in Higher Education, 4* (4), 531-548.

Walters, Kerry, S. (1990). How critical is critical thinking? *Clearing House, 64* (1), 57-60.

Warnick, Barbara & Inch, Edward. (1994). *Critical thinking and communication*, (2nd ed.). New York, McMillan.

Watson, G & Glaser, E. M. (1964). *Critical thinking appraisal manual*. New York: the psychological corporation, Harcourt Brace Jovanovich, In.

Williams Robert L. & Stockdale, Susan. (2003). High-performing students with low critical thinking skills. *The Journal of General Education, 52*(3), 200-226.

Wise, Mary (Executive Director). (2004). *General semantics as critical thinking*: Personal view. International listening Association, center for information & communication science, Ball state University, Muncie, In 47306, ETC.

Wood, Philip & Conner, Jeremey. (1999). Assessment measures. *Assessments Up date, 11*(4). 8-10.

Zimowski, Michele F. Maraki, Eiji, Mislevy, Robert, J & Back, Darrel. (1996). *Bilog- MG: Multiple- Group IRT analysis & test Maintenance for binary items*. Chicago: scientific software.

ملحق (1)

المحكمون

الرقم	اسم المحكم	التخصص	الدرجة العلمية	الوظيفة
.1	عدنان العتوم	علم النفس التربوي	دكتوراه	عميد كلية التربية في جامعة اليرموك
.2	علي الخريشة	مناهج وأساليب تدريس الاجتماعيات	دكتوراه	أستاذ مشارك في جامعة اليرموك
.3	عبد الله الكيلاني	قياس وتقويم	دكتوراه	أ.د. في جامعة عمان العربية
.4	رافع زغول	علم النفس	دكتوراه	أستاذ مشارك في جامعة مؤتة
.5	ضرار جرادات	قياس تقويم	دكتوراه	أ.د. في جامعة اليرموك
.6	عبد الناصر الجراح	علم النفس	دكتوراه	عضو هيئة تدريس في جامعة اليرموك
.7	عبد الله منيزل	قياس وتقويم	دكتوراه	عميد شؤون الطلبة في الجامعة الأردنية
.8	أمل خصاونة	مناهج وأساليب تدريس الرياضيات	دكتوراه	أستاذ مشارك في جامعة اليرموك
.9	إبراهيم رواشدة	مناهج وأساليب تدريس العلوم	دكتوراه	أستاذ مشارك في جامعة اليرموك
.10	عبد الله الصمادي	قياس وتقويم	دكتوراه	أستاذ مشارك في جامعة مؤتة
.11	محمد فريوان	إحصاء	دكتوراه	أ.د. في جامعة اليرموك
.12	موفق حجه	رياضيات	دكتوراه	أ.د. في جامعة اليرموك
.13	شفيق علاونة	علم النفس التربوي	دكتوراه	أ.د. في جامعة اليرموك
.14	عاطف كنعان	لغة عربية	دكتوراه	عضو هيئة تدريس في جامعة البتراء ومستشار تعليم الموهوبين والمتفوقين
.15	يحيى الصمادي	قياس وتقويم	دكتوراه	أستاذ مشارك في الجامعة الأردنية

16. فتحي جروان	علم النفس التربوي	دكتوراه	رئيس قسم علم النفس في جامعة عمان العربية ورئيس المجلس العربي للموهوبين والمتفوقين سابقاً
17. محمد الطيطي	مناهج وأساليب تدريس الاجتماعيات	دكتوراه	مشرف في مدارس الكلية العلمية
18. محمد الدلابيح	لغة عربية	دكتوراه	مشرف لغة عربية في تربية جرش
19. محمد العتوم	قياس وتقويم	دكتوراه	مشرف فيزياء في تربية جرش
20. سمير الرشيد	مناهج وأساليب تدريس الرياضيات	دكتوراه	عضو هيئة تدريس في جامعة جرش
21. أحمد عياصرة	مناهج وأساليب تدريس العلوم	دكتوراه	عضو هيئة تدريس في جامعة جرش
22. حسان بني سلمان	مناهج وأساليب تدريس العلوم	ماجستير	مشرف أحياء في تربية جرش
23. مصطفى جمعة	مناهج وأساليب تدريس الرياضيات	ماجستير	رئيس قسم الإشراف في تربية اربد الأولى
24. محمد الصمادي	مناهج وأساليب تدريس الرياضيات	ماجستير	مدير الشؤون الفنية والتعليمية في تربية اربد الأولى
25. فاضل شطناوي	مناهج وأساليب تدريس الرياضيات	ماجستير	محاضر في جامعة اليرموك
26. علي الشبلي	قياس وتقويم	ماجستير	مشرف رياضيات في تربية جرش
27. منتهى عياش	فيزياء	بكالوريوس	معلمة فيزياء في مدرسة طبريا

ملحق (2)

تعليمات الاختبار

يتضمن هذا الاختبار (82) فقرة تقيس مهارات التفكير الناقد الذي أصبح يعتبر ضرورة من ضرورات الحياة المعاصرة. وتتراوح هذه الفقرات من تلك التي تتطلب تحليل المعنى المتضمن فيها إلى تلك التي تتطلب مهارات معقدة في التفكير الناقد. فبعض الفقرات تتطلب الاستدلال على الاستنتاج الصحيح من خلال مجموعة من الافتراضات، وبعضها يتطلب تقييم الاستنتاج المتوفر، وبعضها الآخر يتطلب الاستنتاج.

- يرجى الإجابة عن جميع الأسئلة.
- لا تضع أي إشارة على كراسة الاختبار.
- ضع إجابتك على ورقة الإجابة المنفصلة المعطاة لك.
- اقرأ كل سؤال بعناية، ثم اختر الإجابة الأفضل من بين البدائل الموجودة لكل سؤال.
- عند تغيير الإجابة تأكد من أنك محوت الإجابة السابقة.
- لكل فقرة أربعة بدائل واحدة منها فقط صحيحة. كما في المثال الآتي:
- الزمن المخصص للاختبار 110 دقائق.

<u>مثال:</u>

حضر ألف طالب من طلاب الصف الثامن اجتماعاً اختيارياً في نهاية الأسبوع في إحدى المدن، واختار الطلاب في هذا الاجتماع موضوعات حول العلاقات بين الأجناس، ووسائل تحقيق سلام عالمي دائم، ليناقشوها، لأنهم شعروا بأنها من أكثر الموضوعات أهمية في الوقت الحاضر.

إحدى العبارات الآتية تمثل الاستنتاج الأقوى من النص السابق.

أ. يتراوح عمر معظم هؤلاء الطلاب ما بين 17-18 سنة.

ب. جاء هؤلاء الطلاب من جميع أنحاء المملكة.

ج. ناقش الطلاب مشكلات تتصل بالعلاقات بين العمال فقط.

د.شعر بعض طلاب الصف الثامن أن مناقشة العلاقات العنصرية ووسائل تحقيق السلام العالمي عمل هام ومفيد.

في المثال السابق، نجد أن الاستنتاج الأول (أ) يحتمل أن يكون خاطئاً لأن المعلومات العامة تدلنا على أن عدداً قليلاً من طلاب الصف الثامن تتراوح أعمارهم بين (17-18) سنة.

وليس هناك دليل يدعم الاستنتاج الثاني (ب) إذ أن البيانات الموجودة ناقصة للحكم على هذه المسألة. والاستنتاج الثالث (ج) خطأ تماماً لأن الحقائق المعطاة تبين أن التلاميذ ناقشوا مشكلات العلاقات العنصرية ووسائل تحقيق السلام العالمي. والاستنتاج الرابع (د) يترتب بالضرورة على الحقائق المعطاة وعلى هذا فهو صحيح تماماً.

من هنا يقوم الطالب باختيار البديل (د) وتظليله كما في الشكل

الفقرة	أ	ب	ج	د
مثال	○	○	○	●

ملحق (3)

عينة من فقرات الاختبار بصورته النهائية (82 فقرة)*

مهارة التحليل:

1. **من اليسير ملء بئر صغير، ومن السهل ملء مخالب فأر. إحدى العبارات الآتية تكمل الجملة السابقة بشكل يتسق معها في المعنى:**

أ. من السهل الانحطاط إلى المنزلة الوضيعة.

ب. من الصعب الارتفاع إلى قمة الجبل.

ج. من الهين إرضاء رجل قنوع.

د. من الصعب النجاح دون تعب.

2. **أعلن الأطباء الألمان أنهم تمكنوا من إجراء عملية زرع للعديد من الأعضاء دفعة واحدة ولأول مرة بأوروبا في تشرين ثاني 2004، حيث تمت زراعة كبدٍ جديدٍ، وبنكرياسٍ ومعدة، واثني عشر، وأمعاء دقيقة، وجزء من الأمعاء الغليظة، في عملية جراحية استغرقت 12 ساعة وقد أعلن الأطباء الأربعة الذين أجروا العملية أن المريضة تتعافى بشكل جيد وأنها تتبع نظاماً غذائياً عادياً.**

يستنتج من الفقرة السابقة أن:

أ. الأطباء الألمان يتفوقون على غيرهم من الأطباء في العالم.

ب. عملية زراعة الأعضاء هي إجراء قديم والأصالة في هذه العملية تتمثل في زراعة عدة أعضاء دفعة واحدة.

ج. عملية زراعة الأعضاء لم تكن معروفة من قبل في ألمانيا.

د. عملية الزراعة للعديد من الأعضاء في العالم معروفة سابقاً والجديد الذي جاءت به ألمانيا هو زراعة هذا النوع من الأعضاء.

---

* يطلب الاختبار من الباحثة مباشرة، أو من قسم الإرشاد وعلم النفس التربوي، جامعة اليرموك.

مهارة التفسير:

3. ترتبط العولمة بخمسة مجالات رئيسة على الأقل، غالباً ما تتجاهل كل منها الأخرى وتشمل: المنافسة الاقتصادية والنقد الاجتماعي وعجز الدولة والثقافة والبيئة الكونية.

أي العبارات الآتية تعبّر عن نفس المعنى؟

أ. ترتبط العولمة بخمسة مجالات رئيسة على الأقل تميل إلى الاستقلال.

ب. ترتبط العولمة بخمسة مجالات رئيسة على الأقل تميل إلى الارتباط القوي الإيجابي.

ج. ترتبط العولمة بخمسة مجالات رئيسة على الأقل تميل إلى الارتباط القوي السلبي.

د. ترتبط العولمة بمجال واحد من بين خمسة مجالات.

4. "إن كنت ريحاً فقد لاقيت إعصاراً" يضرب هذا المثل بالإنسان:

أ. السيئ تكون عواقبه وخيمة.

ب. المتسرع تكون عواقبه وخيمة.

ج. المسالم الهادئ لا يأمن شر المتجبرين من الناس.

د. المداهي الشديد يبلى بمن هو أدهى منه وأشد.

مهارة التقويم:

5. طُلب من صالح أن يحاضر في مجموعة من الشباب غير المبصرين ويختار الموضوع الذي يريد، اختار صالح موضوع عن نعمة البصر، وأهميتها في حياة الإنسان. يتسم الموضوع الذي اختاره صالح بأنه:

أ. مناسب، لأن غير المبصر أكثر الناس حاجة لمعرفة ماذا فقد.

ب. مناسب، لأن غير المبصر أكثر الناس تقديراً لتلك النعمة.

ج. غير مناسب، لأن فاقد الشيء لا يحب أن يذكره أحد بما فقد.

د. غير مناسب، لأن هناك مواضيع أكثر أهمية لغير المبصرين.

6. **يعتبر الإنسان عنصر باني أكثر منه هادم لهذا الكون، إحدى العبارات الآتية تمثل التبرير الأقوى لهذا النص:**

أ. لأن الإنسان أقام العمران والسدود والطرق وشيد الحضارات.

ب. لأن عدد البيوت التي بناها الإنسان أكثر من عدد البيوت التي هدمها.

ج. لان الإنسان لم يهدم الآثار الموجودة منذ العصور القديمة.

د. لأنه تميز ببناء عمارات شاهقة الارتفاع وجسور معلقة.

مهارة الشرح:

7. **يقول الفيلسوف كير كيغارد: "إن المرحلة الأخلاقية غير كافية إذ إن رجل الواجب يمل الاهتمام بواجبه، ومن عدم انتهاك القواعد الحياتية التي وضعها لنفسه، وهو ملل يصل إليه معظم البالغين، فيعودون من جديد إلي المرحلة الجمالية، حيث الحياة أشبه بلعبة، بينما يتجاوز آخرون المرحلة الأخلاقية إلى المرحلة الدينية، ويتجرأون على القفزة الكبيرة إلى أعماق الإيمان، مفضلين الإيمان على متع الحواس وإتمام الواجب الذي يمليه العقل".**

يستدل من النص السابق أن المرحلة الجمالية تعني أن يعيش الإنسان في:

أ. أعماق الإيمان.

ب. تنبه دائم لواجبه.

ج. عالم الحواس.

د. إتمام الواجب الذي يمليه عليه العقل.

8. **إيمان معلمة رياضيات في إحدى المدارس الأساسية للصف الخامس، وقد استخدمت طريقة الألعاب والأنشطة في تدريسها لإحدى الشعب، بينما استخدمت الطريقة التقليدية مع الشعب الأخرى، وذلك لبيان أثر طريقة الألعاب والأنشطة على التحصيل في الرياضيات، وقد وجدت إيمان أن أداء طالبات الشعبة التي استخدمت الألعاب والأنشطة على الاختبار النهائي أفضل من أداء الطالبات في الشعب الأخرى. يمكن الاستدلال من هذه الدراسة ما يأتي:**

أ. يستطيع الطلبة تعلم جميع المواد الدراسية عن طريق الألعاب والأنشطة بصورة أفضل من الطريقة التقليدية.

ب. يفضل جميع الطلبة طريقة الألعاب والأنشطة على الطريقة العادية في التدريس.

ج. إذا أعيدت هذه الدراسة من قبل معلمين ومعلمات آخرين فلن يتوصل أي منهم إلى نفس النتيجة.

د. ستشجع هذه النتيجة المعلمة إيمان على استخدام الألعاب والأنشطة عند توفرها في عملية التدريس.

مهارة الاستدلال:

9. إذا كان فؤاد لا يقول إلا الصدق دائماً، وسامي لا يقول إلا الكذب دائماً، وسألنا فؤاد: إذا سألنا سامي هل سرقت المحفظة؟ فماذا سيجيب؟ أجاب فؤاد: سامي سيقول أنه لم يسرقها. نستدل من الحوار أعلاه أن:

أ. سامي لم يسرق. ب. سامي سرق.

ج. قد يكون سامي سرق. د. البيانات غير كافية للوصول إلى الحل.

10. جميع طلبة المدرسة (س) يتمتعون بذكاء عالٍ، الكثير من الناس في المدينة التي فيها المدرسة (س) يتمتعون بذكاء عالٍ أيضاً.

سناء طالبة في الصف الأول الثانوي وتسكن نفس المدينة.

إحدى العبارات الآتية تمثل استدلالاً صحيحاً من النص السابق

أ. سناء طالبة في المدرسة (س) لكنها ليست ذكية.

ب. سناء طالبة ذكية لكنها ليست في المدرسة (س).

ج. سناء طالبة في المدرسة (س) وذكية.

د. من الصعب أن نحدد وضع سناء سواء من حيث المدرسة أو مستوى الذكاء.

مهارة التبرير الاحتمالي والإحصائي:

11. الشكل أدناه يمثل عدد الأشجار في كل من بستان أياد ولؤي:

بستان أياد	⇧	⇧	⇧	⇧	⇧	⇧	⇧
بستان لؤي	⇧	⇧	⇧	⇧	⇧	⇧	

يمكن أن يكون عدد الأشجار في بستان إياد ولؤي على التوالي ما يأتي:

أ. 70، 60 ب. 70، 50

ج. 56، 44 د. 140، 105

12. **أخذ أحمد 3/1 ما في علبة الشوكولاته، وأخذت نور 2/1 ما تبقى، ثم جاء لؤي وأخذ الباقي:**

تتسم كمية الشوكولاته التي أخذها الأولاد الثلاثة بأنها:

أ. مرتبة تصاعدياً حسب ترتيب أخذهم للشوكولاته.

ب. مرتبة تنازلياً حسب ترتيب أخذهم للشوكولاته.

ج. حصص الأولاد الثلاثة متساوية.

د. الحصة الأكبر لنور يليها أحمد ثم لؤي.

مهارة التعرّف على الافتراضات المنصوص وغير المنصوص عليها:

13. **"لا يمكن أن ينتخب سعيد عضواً في مجلس إدارة النادي هذه الدورة، لأن حق الترشيح يقتصر على أعضاء الجمعية العمومية للنادي دون غيرهم" تستند هذه الحجة على التسليم بأن سعيداً:**

أ. يعتبر من الأعضاء القدامى في النادي.

ب. يرغب في أن يصبح عضواً في مجلس الإدارة.

ج. سيرشح نفسه لمجلس الإدارة ولكنه لن ينجح.

د. ليس عضواً في الجمعية العمومية للنادي.

14. في اجتماع مدير إحدى المدارس مع معلمي المدرسة في بداية العام الدراسي قال:

"دعونا في الحال نتبنى أسلوب التدريس باستخدام الحاسوب، وبذلك نحافظ على مستوى التدريس في مدرستنا".

تقوم هذه الدعوى على إحدى الافتراضات الآتية:

أ. مستوى التدريس في المدارس التي تستخدم الحاسوب أعلى منه في المدارس التي لا تستخدم الحاسوب.

ب. بدأ ينخفض مستوى التدريس في المدرسة عن المستوى المطلوب.

ج. المدير غير راض عن مستوى التدريس الموجود في المدرسة.

د. التدريس في المدرسة لغاية الآن يحقق المستوى المطلوب.

مهارة التركيز:

15. إذا علمت أن س عدد صحيح موجب أقل من 11 ما قيمة حاصل الضرب:

(س− 1) (س− 2) (س− 3) (س − 4) .... (س − 10):

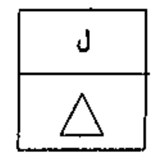
أ. أكبر من صفر.

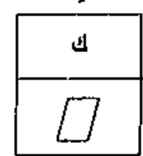
ب. أصغر من صفر.

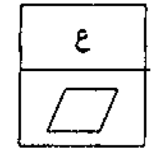
ج. يساوي صفر.

د. س (−1 + 2 −3 +4 −........ +10).

16. إحدى العبارات التالية هي الأكثر دقة في تحديد القاعدة التي كتبت فيها البطاقات الثلاث:







أ. يظهر في البطاقة حرف وشكل هندسي.

ب. إذا كان هناك حرفاً في أحد الجزئين ففي الآخر شكلاً هندسياً.

ج. إذا كان في الجزء العلوي حرفاً فالجزء الأسفل يحوي شكلاً هندسياً.

د. جميع ما ذكر صحيح.

بسم الله الرحمن الرحيم

ملحق 4

جامعـــــة اليرموك

كليـــــــة التربية

قسم الإرشاد وعلم النفس التربوي

قياس وتقويـــــــم

الأستاذ/ الدكتور.. المحترم

السلام عليكم ورحمة الله وبركاته، وبعد:

تقوم الباحثة ببناء اختبار في مهارات التفكير الناقد وذلك استكمالاً لمتطلبات الحصول على درجة الدكتوراه في علم النفس التربوي/ قياس وتقويم في جامعة اليرموك.

ولما عُرف عنكم من خبرة ومعرفة في هذا المجال، فإن الباحثة تأمل أن تستفيد من آرائكم في تطوير هذا الاختبار. راجية إبداء رأيك من خلال إجابتك عن فقرات هذا الاستبيان المكوّن من قسمين.

أشكر لكم جهدكم وتعاونكم.

الباحثة

آمال أحمد الزعبي

فقرات الاستبيان

القسم الأول:

يتعلق بالمهارات التي تم اعتمادها لبناء الاختبار، المطلوب هنا التقييم بنعم أو لا، لمدى مطابقة المهارات مع الفقرة. مع الرجاء بإضافة التعديل أو الاقتراح المناسب في خانة الملاحظات أو المكان المخصص لها.

يشتمل القسم الأول على الأسئلة التالية:

1. هل هناك مهارات أخرى تنطوي تحت مظلة التفكير الناقد وتنسجم مع التعريف غير مذكورة؟ أرجو ذكرها؟
2. هل المهارة منسجمة مع التعريف؟
3. هل المهارات الفرعية التابعة للمهارة الواحدة منسجمة مع هذه المهارة؟
4. هل المهارات الفرعية لنفس المهارة منسجمة مع بعضها؟

القسم الثاني:

يتعلق بفقرات الاختبار: المطلوب هو التقييم بنعم أو لا أمام كل فقرة على ما جاء في كل سؤال من أسئلة هذا القسم، مع الرجاء بإضافة التعديل أو الاقتراح المناسب في المكان المخصص.

يشتمل هذا القسم على الأسئلة التالية:

أولاً: هل الفقرة تمثل المهارة التي اندرجت تحتها.

ثانياً: هل بدائل الفقرة مناسبة.

ثالثاً: هل البديل الصحيح مناسب.

رابعاً: وضوح الفقرات والدقة في صياغتها لغوياً.

نموذج إجابات المحكمين على فقرات القسم الأول من الاستبيان

نموذج إجابة المحكمين للسؤال الأول في القسم الأول:

لا	نعم	ملاحظات (أي مهارات تريد إضافتها)

نموذج إجابة المحكمين للسؤال الثاني في القسم الأول:

المهارة	نعم	لا	ملاحظات
الاستقصاء وحل المشكلات			
معالجة البيانات وتحليلها			
التقويم واتخاذ القرارات			
التفكير الإحصائي والاحتمالي			
التركيز			
الاستدلال			

نموذج إجابة المحكمين للسؤالين الثالث والرابع في القسم الأول:

المهارة	المهارة الفرعية	نعم		لا		ملاحظات
		3	4	3	4	

نموذج إجابة المحكمين بأسئلة القسم الثاني:

الفقرة	نعم				لا				ملاحظات			
	1	2	3	4	1	2	3	4	1	2	3	4
1.												
2.												
3.												
4.												
5.												
6.												
7.												
8.												
9.												
10.												
11.												
12.												
13.												
14.												
15.												
16.												
17.												
18.												
19.												
20.												
21.												
22.												
23.												
24.												
25.												
26.												
27.												
28.												
29.												
30.												

الفقرة	نعم				لا				ملاحظات			
	1	2	3	4	1	2	3	4	1	2	3	4
31.												
32.												
33.												
34.												
35.												
36.												
37.												
38.												
39.												
40.												
41.												
42.												
43.												
44.												
45.												
46.												
47.												
48.												
49.												
50.												
51.												
52.												
53.												
54.												
55.												
56.												
57.												
58.												
59.												
60.												
61.												

الفقرة	نعم				لا				ملاحظات			
	1	2	3	4	1	2	3	4	1	2	3	4
.62												
.63												
.64												
.65												
.66												
.67												
.68												
.69												
.70												
.71												
.72												
.73												
.74												
.75												
.76												
.77												
.78												
.79												
.80												
.81												
.82												
.83												
.84												
.85												

ملحق 5

القيمة العظمى لدالة المعلومات للفقرات وقيمة القدرة المقابلة للقيمة العظمى لدالة المعلومات

الفقرة	القيمة العظمى لدالة المعلومات	قيمة القدرة المقابلة للقيمة العظمى لدالة المعلومات
1	0.12	0.45
2	1.06	1.22
3	0.53	0.93
4	0.23	0.65
5	0.17	-0.92
6	0.19	-0.97
7	1.33	1.41
8	0.28	0.98
9	1.30	1.11
10	0.07	-0.45
11	0.58	1.14
12	0.79	0.88
13	1.06	1.28
14	1.93	1.01
15	0.11	-0.26
16	0.97	1.32
17	0.21	0.95
18	0.38	0.92
19	0.12	-0.89
20	0.88	1.33
21	0.43	0.80
22	0.76	1.22
23	0.27	0.71
24	0.13	0.31
25	2.04	1.71
26	0.31	0.53
27	0.98	1.24
28	1.78	1.52
29	1.13	1.35
31	1.98	1.35
32	1.02	0.89
33	0.16	0.49
34	0.48	1.47
35	0.63	1.43
36	0.08	0.29
37	0.17	-0.24
38	0.22	0.46
39	0.30	1.43
40	0.54	1.53
41	0.24	0.73
42	0.34	1.13
43	0.52	0.66
44	0.50	1.26
45	0.48	1.21
46	0.23	0.94
47	0.11	0.61
48	0.49	0.40
49	0.39	1.15
50	1.22	1.28
51	0.56	0.81
53	0.28	1.52
55	1.07	1.73
56	0.49	1.15
57	0.08	0.63
58	0.14	0.99
60	0.37	1.10
61	0.05	0.74
62	1.14	1.56
63	0.41	1.59
64	0.28	0.98
65	0.18	0.14
66	0.20	0.70
67	0.14	0.19
68	0.11	0.26
69	1.18	1.14
71	1.27	1.52
72	0.57	-0.51
74	0.20	1.32
75	0.20	0.51
76	0.26	0.57
77	0.53	0.26
78	0.07	0.61
79	0.38	0.38
80	0.28	0.83
81	0.81	1.23
82	0.34	0.52
84	0.34	0.65
85	0.28	1.41
86	1.97	1.67
87	0.22	-0.05
88	0.29	-0.34
89	0.45	0.23

Abstract

Al-Zoubi, Amal Ahmad. Constructing a Critical Thinking Skills Test by Using Item Response Theory. Ph.D. Dissertation, Yarmouk University, 2006 (Supervisor: Prof. Yousef Mohammad Sawalmeh).

This study aimed at constructing a critical thinking skills test by using item response theory. To achieve the purposes of the study the concept of critical thinking was defined, the critical thinking skills were derived, and 228 four - choice items were written to measure the following skills: Analysis, interpretation, evaluation, explanation, inference, focusing, recognition of stated and unstated assumptions, and statistical and probabilitical reasoning. Eighty nine items were retained as a result of judgment process and piloting.

The test was administrated to 444 students in high primary school, 610 students in high secondary school, 120 university students. The reliability, validity and unidimensionlaity of the test was established by many different indices.

Regarding the model goodness of fit, 50 items were appropriate to 1PL, 70 items were appropriate to 2 PL, and 82 items were appropriate to 3 PL.

SPSS & Bilog- Mg were used to analyze the data, through this process estimates of persons abilities, items parameters, relative precision and relative efficiency were computed. The

findings of the analysis indicated that the test according to 3PL model was the best among the other models especially at the high ability levels. The findings also indicated that short version tests according to all models (1 PL, 2PL, 3 PL) composed of easy items were appropriate to low ability individuals, with the test according to the 3 PL was the best.

The invariance of item parameters and persons abilities was investigated, the results showed that the invariance of item parameters & persons abilities were approximately convenient in 67% cases for 1PL, in 67% cases for 2PL and in 78% cases for 3PL. Percentiles of the total score for different norm groups were computed.

As result of that, the researcher recommended using the test or any short version test to measure the critical thinking skills also using other format to develop a test of critical thinking by using, item response theory.

Key words:

Critical Thinking, Item Response Theory, Relative Precision, Relative Efficiency, Invariance.